MASS SPECTROMETRY BASED UNTARGETED METABOLOMICS IN FOOD AND HEALTH AREA

María Leticia Lacalle Bergeron

February 2022

Supervisors:

Prof. Dr. Juan V. Sancho Llopis

Dr. Tania Portolés Nicolau

UNIVERSITAT JAUME·I

UNIVERSITAT JAUME·I

Doctoral Programme in Sciences

Doctoral School of Universitat Jaume I

MASS SPECTROMETRY BASED UNTARGETED METABOLOMICS IN FOOD AND HEALTH AREA

Thesis presented by María Leticia Lacalle Bergeron in fulfilment of the requirements for the degree of Doctor (PhD) from Universitat Jaume I

MARIA LETICIA\| LACALLE\| BERGERON

Firmado digitalmente por MARIA LETICIA\| LACALLE\|BERGERON
Fecha: 2022.02.07 17:08:51 +01'00'

María Leticia Lacalle Bergeron

TANIA\| PORTOLES\| NICOLAU

Firmado digitalmente por TANIA\|PORTOLES\| NICOLAU
Fecha: 2022.02.07 17:50:04 +01'00'

Dr. Tania Portolés Nicolau

JUAN VICENTE\| SANCHO\|LLOPIS

Firmado digitalmente por JUAN VICENTE\| SANCHO\|LLOPIS
Fecha: 2022.02.07 17:14:04 +01'00'

Prof. Dr. Juan V. Sancho Llopis

Castellón de la Plana, February 2022

Funding received

This work responds to the commitment acquired with Universitat Jaume I in Spain, for granting María Leticia Lacalle Bergeron with a predoctoral contract belonging to the *Plan de Promoción de la Investigación* of Universitat Jaume I for the training of research personnel, since 1st October 2019 (UJI 19I001/03).

María Leticia Lacalle Bergeron has also received a grant for a research stay in *Service de Pharmacologie et Immunoanalyse* (SPI), *Laboratoire d'Etudes du Métabolisme de Médicaments* (LEEM) from the French Alternative Energies and Atomic Energy Commission (CEA), in France from 17th of December 2018 to 31st of March 2019; under the supervision of Dr. Christophe Junot et Dr. François Fenaille. This stay was funded by the *Fundación Balaguer Gonel Hermanos* in Spain and belonging to the *Convocatoria de ayudas, con financiación de la Fundación Balaguer Gonel Hermanos, para estancias de investigación del alumnado de la Escuela de Doctorado de la Universitat Jaume I que realiza tesis doctorales que opten a la mención de doctor o doctora internacional.*

The work included in this thesis have been funded by *Generalitat Valenciana* (PROMETEO/2019/040) and Universitat Jaume I (UJI-B2020-25 and UJI-B2020-37).

Thesis by compendium of publications:

Scientific article I. L. Lacalle-Bergeron, D. Izquierdo-Sandoval, J.V. Sancho, F.J. López, F. Hernández and T. Portolés. Chromatography hyphenated to high resolution mass spectrometry in untargeted metabolomics for investigation of food (bio)markers. Trends in Analytical Chemistry 2021; 135:116161. doi: 10.1016/j.trac.2020.116161. Impact factor 12.296 (2020).

Scientific article II. L. Lacalle-Bergeron, T. Portolés, C. Sales, M.C. Corell, F. Domínguez, J. Beltrán, J.V. Sancho and F. Hernández. Gas chromatography-mass spectrometry based untargeted volatolomics for smoked seafood classification. Food Research International 2020, 137:109698. doi: 10.1016/j.foodres.2020.109698. Impact factor 6.475 (2020).

Scientific article III. L. Lacalle-Bergeron, T. Portolés, F.J. López, J.V. Sancho, C. Ortega-Azorín, E.M. Asensio, O. Coltell and D. Corella. Ultra-performance liquid chromatography-ion mobility separation-quadrupole time-of-flight MS (UHPLC-IMS-QTOF MS) metabolomics for short- term biomarker discovery of orange intake: a randomized, controlled crossover study. Nutrients 2020; 12:1916. doi: 10.3390/nu12071916. Impact factor 5.719 (2020).

Scientific article IV. L. Lacalle-Bergeron, D. Izquierdo-Sandoval, A. Fernández-Quintela, M. Puy Portillo, J.V. Sancho, F. Hernández and T. Portolés. The potential of ion mobility separation in combination with high resolution mass spectrometry for the identification of biomarkers highlighted by untargeted metabolomics: the effects of pterostilbene and resveratrol consumption in liver steatosis, animal model. Food Chemistry (2021) Submitted. Impact factor 7.514 (2020).

Scientific article V. L. Lacalle-Bergeron, R. Goterris-Cerisuelo, T. Portolés, J. Beltrán, J.V. Sancho, C. Navarro-Moreno and F. Martinez-Garcia. Novel sampling strategy for alive animal volatolome extraction combined with GC-MS based untargeted metabolomics: identifying mouse pup pheromones. Talanta 2021; 235:122786. doi: 10.1016/j.talanta.2021122786. Impact factor 6.057 (2020).

Scientific article VI. L. Lacalle-Bergeron, R. Goterris-Cerisuelo, J. Beltrán, J.V. Sancho, C. Navarro-Moreno, F. Martinez-Garcia and T. Portolés. Untargeted metabolomics approach to putative pheromones in mice. Part 2: using UHPLC-IMS-QTOF MS for surface body samples to identify low-volatility chemosignals. Talanta (2021) Submitted. Impact factor 6.057 (2020).

"This thesis has been accepted by the co-authors of the publications listed above that have waved the right to present them as a part of another PhD thesis"

This thesis has been developed and will be defended according to the requirements for obtaining the International PhD degree:

1. The present thesis has been written in English. The summary, objectives and conclusions and future work sections have also been included in Spanish and French.

2. The candidate has performed a research stay in *Service de Pharmacologie et Immunoanalyse* (SPI), *Laboratoire d'Etudes du Métabolisme de Médicaments* (LEEM) from the French Alternative Energies and Atomic Energy Commission, hereinafter referred to as CEA, in Saclay (France); under the supervision of Dr. Christophe Junot and Dr. François Fenaille.

3. Prior to the defense, this work has been evaluated by two international independent reviewers directly related with the research field, Prof. Dr. Serge Rudaz (University of Geneva, Switzerland) and Dr. Josep Rubert (Wageningen University & Research, Netherlands). Additionally, three national independent reviewers have also evaluated this work, Prof. Dr. Coral Barbas Arribas (Universidad CEU San Pablo, Spain), Dr. Maykel Hernández Mesa (Universidad de Granada, Spain) and Dr. María Ibáñez Martínez (Universitat Jaume I, Spain).

4. The thesis will be defended in Spanish and English, and at least one international expert from a non-Spanish university will be part of the tribunal panel.

A mi familia et ma famille, y en especial a mi queridísimo padrino.

Agradecimientos/Remerciements/Acknowledgements

Después de estos 5 años, la tesis está acabada. Parece que fue ayer cuando emprendí este viaje, y a la vez, siento que han pasado mucho tiempo. A esa chica que todavía en tercero no tenía claro si se había metido en la carrera que tocaba, le parecería increíble haber llegado hasta aquí. A la de ahora todavía le cuesta asimilarlo. No fue hasta que tuve la suerte de hacer unas prácticas en el departamento de química analítica del CIT en el año de Erasmus, ya de por si un año revolucionario, que me dije a mi misma "vale… ahora sí… ahora sé por dónde ir". Como cambia tu mentalidad cuando tienes claro lo que quieres hacer. Y aun en ese momento solo me sentía segura sobre la rama de la química que me gustaba, sin imaginarme que le dedicaría 6 años más de estudios. Ha sido una etapa de gran crecimiento profesional y también personal, que no sería la misma sin todas esas personas que me han acompañado.

En primer lugar, querría agradecer la inestimable ayuda y dirección de la *Dra. Tania Portolés* y del *Dr, Juanvi Sancho*, sin los cuales esta tesis no habría sido posible. A *Tania*, por ser una guía y un apoyo en todo momento, a la que he acudido tanto en momentos de crisis como de alegría. Por empujarme en los momentos duros cuanto tú tenías más confianza en mí que yo misma. Por darme ánimos y enseñarme a perseverar cuando yo quería rendirme. Por esos momentos de risa y todas esas conversaciones, tertulias, *brainstormings*; tanto en el café como en el despacho, que versaban tanto de trabajo como de nuestras pequeñas grandes obsesiones y hobbies; o simplemente sobre nosotras. A *Juanvi*, porque desde que pise el IUPA como estudiante de TFG me ha mostrado el mundo de la investigación. Por contagiar su entusiasmo, por tener siempre abierta la puerta del despacho a cualquier duda (de cualquier tema), por su buen humor, por ser proveedor de cualquier capítulo de serie o película, por las comidas en el despacho viendo GoT y las charlas en el café. Por todo, gracias.

Al *Dr. Félix Hernández* querría agradecerle la oportunidad de formar parte del IUPA e introducirme en la investigación, por transmitirnos sus consejos y conocimiento.

También querría mencionar a todos los co-autores de los trabajos que he ido llevando a lo largo de estos años. Esta tesis no se podría haber llevado a cabo sin la estrecha colaboración con grupos de investigación o empresas de distintos ámbitos, recalcando la importancia de la colaboración multidisciplinar. En primer lugar, a Fernando Domínguez y a Mª Carmen Corell de Sea Delight Europe, S.L.. A la Dra. Dolores Corella del Departamento de Medicina Preventiva y Salud Pública, Ciencias de la Alimentación, Toxicología y Medicina Legal de la Universidad de Valencia, junto con Dr. Oscar Coltell, Carolina Ortega Azorín y Eva M. Asensio. A la Dra. María Puy Portillo y el Dr. Alfredo Fernández del Departamento de Nutrición y Ciencia de los Alimentos de la universidad del País Vasco. Por último, pero no menos importante, al Dr. Fernando Martínez García, Rafa Goterris Cerisuelo y Cinta Navarro Moreno de la Unidad Predepartamental de Medicina de la Universidad Jaume I.

Remerciements particuliers au Dr François Fenaille et le Dr Christophe Junot du Service de Pharmacologie et d'Immunoanalyse du CEA Paris-Saclay, pour m'avoir donné l'opportunité et toutes les facilités pour faire le stage pré-doctoral avec vous. Un merci spécial à la Dr Kathleen Rousseau, avec qui j'ai partagé jour après jour de ce séjour. Le temps que j'ai passé avec vous m'a permis d'approfondir les aspects plus théoriques et instrumentaux de la métabolomique avec la spectrométrie de masse d'haute résolution.

No puedo olvidarme de todos los compañeros y profesores que tengo y he tenido en el Instituto Universitario de Plaguicidas y Aguas. *Elena, Roque, Toni, Ana María, Jose, Arantxa, Carmen, Noelia*, por su ayuda durante estos años. *Robert, Edu*, *Clara*, *Inés* y *Jaime* por esos primeros años que empecé con entusiasmo, por las cenas y las cervezas/vinos. A *Montse* y a *Mercedes* por vuestro incondicional apoyo, consejos de amigas, risas y muchos etcs. A *Borja* y a *Jorge*, por esas conversaciones en la salita del café o en becaría y a *Carlos* por ayudarme a pelearme con el GC-2. *Vima*, por ser la sonrisa del IUPA. A *Marina*, por traer tu alegría desde el sur. A *Paco* por sus consejos y apreciar mi ensalada. A *Claudia*, *Elisa* y *María Mata*, que empiezan esta nueva etapa. Al miembro más reciente de IUPOMICS, *David Izquierdo*

(*Mr Leftman*), por el entusiasmo con el que hace todo, desde videollamadas en pleno confinamiento hasta pegarse con el VION.

En especial me gustaría agradecer a Rubén, con el que empecé esta etapa. Por contagiarme su buen humor y entusiasmo, por dedicarme su tiempo. Que aun a distancia está ahí para lo que necesites.

A María por su apoyo, por estar presente desde el principio, por ser compañera de viaje y por las charlas y las risas. Por marcar el comienzo de las vacaciones de verano con paella y piscina en la Vilavella.

A Ximo su alegría, por siempre sacar unos minutos para ayudarme cuando lo necesitaba: cuando no sabíamos por dónde pillar la estadística, cuando al GC-2 le da por hacer de las suyas, y un largo etcétera.

Y por supuesto me gustaría agradecer a dos personas con los que he vivido día a día esta montaña rusa que es la tesis, compañeros incluso antes de empezar el master, a los *Cromatografistas Viajeros*, *David* y *Alberto*. A *David* por estar ahí siempre que lo necesitabas, ayudándote en cualquier cosa que le pidieras, paladín en la luchar contra el VION cuando no le daba la gana funcionar, "legión de reconocimiento" cuando había un misterio que resolver. Compañero de despacho en estos últimos años, compartiendo conversaciones de todo tipo, con un abrazo preparado para cuando lo necesitaras. Y *Alberto*, por tantos momentos, tantas meriendas (¡de media hora, *of course*!), tantas tensiones, alegrías, consejos (de los cuales nuuunca hacemos caso), clases de inglés, tardes en el IUPA… Tantos "*Don't worry, be happy*" ya desde el master, donde apuntábamos a como seriamos durante la tesis, tu nervioso perdido y yo apurando hasta el final. No sé si algún día conseguiré que veas una (¡solo te pido una!) de las pelis que te recomiendo… Muchas gracias.

Y por supuesto, no puedo olvidarme de aquellas personas que me han acompañado a nivel personal.

A mis "chic@s de la UJI" *Marieta*, *Alba Corazón*, *Adrián*, *Amanda* y *Eli*. Por todos esos momentos que hemos vivimos desde que nos conocimos en 1º de grado en química, en la cantina, en clase, y después, sobre todo después.

A mis F.R.I.E.N.D.S, *Eva* y *María*. Por estar ahí desde siempre, por todo lo que hemos pasado juntas, por ser las más payasas, por ser fuente de alegría y de risas que en seguida te quitan todas las tensiones.

A mi familia Lacalle, a todos: primas, primos, titos y titas. A mi *Alula,* fuente de cariño. Porque, aunque no nos veamos muchos, siempre hay cariño y risas. Al tito *Pepito* y mi abuelo *Pepe*.

A ma famille française, toute la famille. Parce que même dans les quelques moments où nous pouvons nous retrouver, il semble que le temps ne s'est pas écoulé. A ma Mutti, qui est toujours là pour aider.

Y por supuesto los más importantes. A papá y mamá, que suerte teneros. No encontraré nunca las palabras adecuadas para agradecer todo lo que habéis hecho por mí, por siempre estar ahí, por el todo el cariño y por todo el apoyo. Por haberme dado los mejores hermanos del mundo, Sofia y Tomás. Porque mi infancia y mi vida no habría sido tan bonita sin ellos. *Je vous aime très fort*.

¡Gracias a todos!

Merci à tous !

Summary

The impressive scientific advances of the last decades has allowed the development of highly advanced analytical methodologies and instrumentation, as well as their application to very diverse fields, which has greatly increased knowledge about biological systems and their functions. Furthermore, the development of bioinformatics tools allowed the handling of complex data matrices that a few decades ago it would have been unimaginable. Since the end of the nineties, metabolomics, the last of the -*omics* sciences in charge of the study of metabolites from a biological system and their changes promoted by endogenous or exogenous conditions, has emerged with force. Specifically, the objective of untargeted metabolomics is to qualitatively or semi-quantitatively analyse the maximum number of compounds as possible and select, using computer and statistical tools, those that truly provides information to the problem under study. The focus in this type of the study is not based on restricted starting hypotheses, but rather these are generated after addressing the problem globally and based on the answers obtained The sensitivity and the ability to cover a large number of metabolites in complex matrices make advanced analytical techniques based on chromatography coupled to mass spectrometry the preferred ones for untargeted metabolomics studies.

This approach has been applied in many research areas related to food, health and the study of biological systems of a very different nature, among others; diffusing the boundaries between different fields of study and creating multidisciplinary areas.

In this doctoral thesis, the contribution of untargeted metabolomics together with advanced analytical techniques based on the coupling of chromatographic separations with both low resolution mass spectrometry (MS) and high resolution mass spectrometry (HRMS) was evaluated in different studies related to the field of food, health and intra-especies chemical communication. In addition, the advantages in the elucidation of unknown compounds provided by the latest generation instruments based on ion mobility separation in combination with high resolution mass spectrometry (IMS-HRMS) were explored.

The studies carried out during the doctoral period have resulted in the publication of 6 scientific articles in peer-reviewed journals, which form the basis of this thesis, made up of 6 differentiated chapters. After the first chapter dedicated to the introduction, two main blocks can be differentiated. The first block was divided into chapters 2, 3 and 4 related to the application of metabolomics in the field of food and its relationship with health. The second block comprises chapter 5 where the potential of untargeted metabolomic approaches in the investigation of intra-species chemical communication was investigated. Finally, in chapter 6 the main conclusions from this PhD work are summarised and the derived further research studies are presented.

The objective of the first chapter is to be presented as an Introduction to offer a broad vision of the topic of this doctoral thesis. It is mainly formed by **Scientific Article I**, where a detailed description of the untargeted metabolomic workflow and the widely employed strategies based on chromatography coupled to mass spectrometry for the identification of (bio) markers in the food field are provided. In addition, a bibliographic review is made on the trends of the last 5 years in each phase of the metabolomic process. Although this review article is focused on the food field, what is described is applicable to different areas, including those addressed in this thesis.

The second chapter is made up of **Scientific Article II**. The objective of this study was to find volatile markers that confirm the difference between smoked fish products with different processing techniques and that successfully concluded in a classification model based on 11 marker compounds from the different confronted fish treatment processes. The volatiles were extracted from the fish by dynamic extraction using purge-and-trap (DHS-P&T) and analysed by GC-MS. This article exemplifies the applicability of untargeted metabolomics on food processing, which includes quality control, food safety and authentication, among others.

In the third chapter, an untargeted metabolomic approach is employed for the identification of food intake biomarkers, exemplified by the discovery of short-term (**Scientific Article III**) and medium-term biomarkers of orange intake. Furthermore, the advantages that the new IMS-HRMS

combination could bring to the discovery of new dietary biomarkers were studied, ending in the tentative identification of 6 short-term biomarkers of orange intake.

Chapter 4 is the last one focused on the food field, specifically in the effects that food has on health. In **Scientific Article IV**, the effect of bioactive compounds such as resveratrol and pterostilbene on hepatic steatosis was studied. For this, an animal model of rats with induced hepatic steatosis by a diet high in fat and fructose was employed. In addition, the different tools emerged after the implementation of IMS-HRMS that facilitate the identification of biomarkers were evaluated, such as CCS prediction tools.

In the second block, chapter 5 is focused on the application of untargeted metabolomics to fields that have not been approached until now with an interdisciplinary vision, such as the discovery of chemical compounds used in communication between individuals of the same species. Specifically, the objective of this chapter was the determination of compounds emitted by mouse pups that induce an activation of the maternal instinct. The main novelty that had to be carried out was in the sampling stage, since not only involved the work with alive animals, but the sampling conditions should not interfere with the development of the pups or induce excessive stress that could affect the study. On the one hand, **Scientific Article V** was focused on compounds with a more volatile nature, adapting the DHS-P&T extraction to capture the volatolome from alive animals, which was subsequently analysed by GC-MS. The identity of 11 volatile compounds from mouse pups putatively involved in the induction of maternal care in adult females was discovered. On the other hand, **Scientific Article VI** was focused on those low-volatility compounds. Samples were taken by gently rubbing swabs on the facial and anogenital areas of the mouse pups. The extracts were analysed by LC-IMS-HRMS and 5 compounds from neonatal mice with a possible role in materno-filial communication were tentatively identified.

Resumen

El espectacular progreso científico de las últimas décadas ha permitido el desarrollo de metodologías e instrumentación analítica altamente avanzadas, así como su aplicación a muy diversos campos que ha incrementado enormemente el conocimiento sobre los sistemas biológicos y sus funciones. Además, el desarrollo de las herramientas bioinformáticas ha permitido abarcar matrices de datos tan complejas que hace unas décadas hubiera sido inimaginable. Desde finales de los noventa, la metabolómica, la última de las ciencias *-ómicas* encargada del estudio del conjunto de metabolitos de un sistema biológico y los cambios promovidos por condiciones endógenas o exógenas, ha emergido con fuerza. Concretamente, la metabolómica no dirigida tiene como objetivo analizar de manera cualitativa o semicuantitativa el máximo número de compuestos posible y seleccionar mediante herramientas informáticas y estadísticas aquellos que verdaderamente aporten información a la problemática bajo estudio. El enfoque de este tipo de estudios no se basa en hipótesis de partida restringidas, sino que estas se generan después de abordar el problema de manera global y en base a las respuestas obtenidas. La sensibilidad y la capacidad de abarcar un gran número de metabolitos en matrices complejas hacen de las técnicas analíticas avanzadas basadas en cromatografía acoplada a espectrometría de masas las preferidas en los estudios de metabolómica no dirigida.

Esta aproximación se ha aplicado en multitud de áreas de investigación relacionadas con la alimentación, la salud y el estudio de sistemas biológicos de naturaleza muy distinta, entre otras, llegando a hacer difusa la frontera entre distintos campos de estudio y creándose áreas multidisciplinares.

En esta tesis doctoral se han evaluado la aportación de la metabolómica no dirigida en combinación con técnicas analíticas avanzadas basadas en el acoplamiento de cromatografía con espectrometría de masas tanto de baja (MS) como de alta resolución (HRMS) en distintos estudios relacionados con el campo de la alimentación, la salud y la comunicación química intraespecie. Además, se han evaluado las ventajas que aportan los instrumentos de última generación basados en separación por movilidad iónica en combinación con

espectrometría de masas de alta resolución (IMS-HRMS) en la elucidación de compuestos desconocidos.

Los estudios realizados durante el periodo doctoral han resultado en la publicación de 6 artículos científicos en revistas revisadas por pares, que conforman la base de esta tesis dividida en 6 capítulos diferenciados. Después del primer capítulo dedicado a la introducción, se pueden diferenciar dos bloques. El primero estaría formado por los capítulos 2, 3 y 4 relacionados con la aplicación de la metabolómica no dirigida en el campo de la alimentación y su relación con la salud. El segundo bloque comprende el capítulo 5 donde se investigó el potencial de las aproximaciones metabolómicas no dirigidas en la investigación de la comunicación química intraespecie. Finalmente, en el capítulo 6 se recogen las conclusiones principales de los trabajos realizados en esta tesis doctoral y los futuros trabajos de investigación derivados.

El objetivo del primer capítulo es presentar como Introducción una visión amplia de la temática objetivo de esta Tesis Doctoral. Lo conforma principalmente el **Artículo científico I**, donde se realiza una descripción detallada del flujo de trabajo de la metabolómica no dirigida y de las estrategias ampliamente empleadas basadas en la cromatografía acoplada a espectrometría de masas para la identificación de (bio)marcadores en el ámbito de la alimentación. Además, se hace una revisión bibliográfica sobre las tendencias de los últimos 5 años en cada fase del proceso metabolómico. A pesar de que este artículo de revisión está focalizado en el campo de la alimentación, lo descrito es aplicable a distintos ámbitos, incluyendo los abordados en esta tesis.

El segundo capítulo se compone del **Artículo científico II**. El objetivo de este estudio fue encontrar marcadores volátiles que confirmaran la diferencia entre productos pesqueros ahumados con diferentes técnicas de procesado y que culminó satisfactoriamente en un modelo de clasificación basado en 11 compuestos marcadores de los distintos procesos de tratamiento de pescado estudiados. Los compuestos volátiles fueron extraídos del pescado mediante extracción dinámica por purga y trampa (DHS-P&T) y analizados por GC-MS. Con este artículo se ejemplifica la aplicabilidad de la metabolómica no dirigida en el procesamiento

alimentario, que incluye control de calidad, seguridad y autentificación de alimentos, entre otros.

En el tercer capítulo se emplea la metabolómica no dirigida para la identificación de marcadores de ingesta de alimentos, ejemplificado con el descubrimiento de marcadores de ingesta de naranja a corto (**Artículo científico III**) y medio plazo. Además, se estudiaron las ventajas que la nueva combinación IMS-HRMS podía aportar para el descubrimiento de nuevos biomarcadores dietéticos, culminando en la identificación tentativa de 6 marcadores de ingesta de naranjas.

El capítulo 4 es el último focalizado en el ámbito de la alimentación, concretamente en el efecto que tiene la alimentación en la salud. En el **Artículo científico IV** se estudió el efecto de compuestos bioactivos como el resveratrol y el pterostilbeno en la esteatosis hepática. Para ello, se empleó un modelo animal de ratas con esteatosis hepática inducida por una dieta alta en grasas y en fructosa. Además, se han evaluado distintas herramientas que están surgiendo para facilitar la identificación de biomarcadores gracias a la implementación de instrumentos IMS-HRMS, como las herramientas de predicción de CCS.

En el segundo bloque, el capítulo 5 se focalizó en la aplicación de la metabolómica no dirigida a campos poco abordados hasta ahora con una visión interdisciplinar, como es el descubrimiento de compuestos químicos empleados en la comunicación entre individuos de la misma especie. Concretamente, el objetivo de este capítulo era la determinación de compuestos emitidos por crías de ratón que indujeran a una activación del instinto maternal. La principal novedad que se tuvo que realizar se centró en la etapa de muestreo, dado que no solo se tenía que trabajar con animales vivos, sino que las condiciones del muestreo no debían interferir en el desarrollo de las crías ni inducir un estrés excesivo que pudiera afectar al estudio. Por una parte, el **Artículo científico V** se centra en compuestos de naturaleza más volátil, adaptando la extracción DHS-P&T para captar el volatoloma de los animales vivos, que posteriormente se analizó por GC-MS. Se determinó la identidad de 11 compuestos volátiles provenientes de crías de ratón putativamente implicados en la inducción del cuidado materno en hembras adultas. Por otra parte, el **Artículo científico VI** se focalizó en

aquellos compuestos de carácter menos volátil. Las muestras se tomaron frotando suavemente hisopos sobre zonas facial y anogenital de las crías de ratón. Los extractos se analizaron por LC-IMS-HRMS y se identificaron tentativamente 5 compuestos procedentes de ratones neonatales con una posible implicación en la comunicación maternofilial.

Résumé

Les progrès scientifiques spectaculaires des dernières décennies ont permis le développement de méthodologies et d'instruments analytiques très avancés, ainsi que leur application dans des domaines très divers qui ont considérablement augmenté les connaissances sur les systèmes biologiques et leurs fonctions. De plus, le développement des outils bioinformatiques a permis de couvrir des matrices de données si complexes qu'il y a quelques décennies cela aurait été inimaginable. Depuis la fin des années 90, la métabolomique, la dernière des sciences *-omiques* chargées de l'étude de l'ensemble des métabolites d'un système biologique et des changements favorisés par les conditions endogènes ou exogènes a émergé avec force. Concrètement, l'objectif de la métabolomique non ciblée est d'analyser qualitativement ou semi-quantitativement le maximum de composés possibles et de sélectionner, à l'aide d'outils informatiques et statistiques, ceux qui apportent véritablement des informations sur la problématique étudiée. L'objectif de ce type étude n'est pas basé sur des hypothèses de départ restreintes, mais plutôt celles-ci sont générées après avoir abordé le problème globalement et sur la base des réponses obtenues. La sensibilité et la capacité de couvrir un grand nombre de métabolites dans des matrices complexes font des techniques analytiques avancées basées sur la chromatographie couplée à la spectrométrie de masse les préférées dans les études de métabolomique non ciblée.

Cette approche a été appliquée dans de nombreux domaines de recherche liés à l'alimentation, à la santé et à l'étude de systèmes biologiques de nature très différente, entre autres ; arrivent à diffuser la frontière entre les différents domaines d'études et créer des secteurs multidisciplinaires.

Dans cette thèse doctorale, la contribution de la métabolomique non ciblée a été évaluée en combinaison avec des techniques analytiques avancées basées sur le couplage de séparations chromatographiques avec la spectrométrie de masse à la fois basse (MS) et haute résolution (HRMS) dans différentes études liées au domaine de l'alimentation, la santé et la communication chimique intraspécifique. De plus, les avantages des instruments de dernière génération basés sur la séparation par mobilité

ionique en combinaison avec la spectrométrie de masse à haute résolution (IMS-HRMS) dans l'élucidation de composés inconnus ont été évalués.

Les études menées au cours de la période doctorale ont conduit à la publication de 6 articles scientifiques dans des journal révisé par des pairs, qui constituent la base de cette thèse divisée en 6 chapitres différenciés. Après le premier chapitre consacré à l'introduction, deux blocs peuvent être différenciés. Le premier serait composé des chapitres 2, 3 et 4 relatifs à l'application de la métabolomique non ciblée dans le domaine de l'alimentation et de ses relations avec la santé. Le deuxième bloc comprend le chapitre 5 où a été étudié le potentiel des approches métabolomiques non ciblée dans l'étude de la communication chimique intraspécifique. Finalement, dans le chapitre 6 sont rassemblées les principales conclusions des travaux menés dans cette Thèse Doctorale et les futurs travaux de recherche qui en découlent.

L'objectif du premier chapitre est de présenter en Introduction une vision large de l'objet de cette Thèse Doctorale. Il est principalement composé de l'**Article scientifique I**, où est faite une description détaillée du flux de travail de la métabolomique non ciblée et des stratégies largement utilisées basées sur la chromatographie couplée à la spectrométrie de masse pour l'identification de (bio)marqueurs dans le domaine de l'alimentation. De plus, une revue bibliographique est indiquée sur les tendances des 5 dernières années dans chaque phase du processus métabolomique. Malgré le fait que cet article de synthèse se concentre sur le domaine de l'alimentation, ce qui est décrit s'applique à différents domaines, y compris ceux abordés dans cette thèse.

Le deuxième chapitre est constitué de l'**Article scientifique II**. L'objectif de cette étude était de trouver des marqueurs volatils qui confirment la différence entre les produits de la pêche fumés avec différentes techniques de traitement et qui ont abouti avec succès à un modèle de classification basé sur 11 composés marqueurs des différents procédés de traitement du poisson étudiés. Les composés volatils ont été extraits des poissons par extraction dynamique de l’espace de tête avec piégeage par

sorbant (DHS-P&T) et analysés par GC-MS. Cet article illustre l'applicabilité de la métabolomique non ciblée dans le traitement alimentaire, qui comprend le contrôle de la qualité, la sécurité alimentaire et l'authentification, entre autres.

Dans le troisième chapitre, la métabolomique non ciblée est utilisée pour l'identification de marqueurs de consommation alimentaire, illustrée par la découverte de marqueurs de consommation d'orange à court (**Article scientifique III**) et à moyen terme. De plus, les avantages que la nouvelle combinaison IMS-HRMS pourrait apporter à la découverte de nouveaux biomarqueurs alimentaires ont été étudiés, aboutissant à l'identification provisoire de 6 marqueurs de la consommation d'orange.

Le chapitre 4 est le dernier consacré au domaine de l'alimentation, en particulier l'effet de l'alimentation sur la santé. Dans l'**Article scientifique IV,** l'effet de composés bioactifs tels que le resvératrol et le ptérostilbène sur la stéatose hépatique a été étudié. Pour cela, un modèle animal de rats atteints de stéatose hépatique induite par un régime riche en graisses et en fructose a été utilisé. De plus, les différents outils qui émergent pour faciliter l'identification de biomarqueurs grâce à la mise en place d'instruments IMS-HRMS ont été évalués, comme les outils de prédiction CCS.

Dans le deuxième bloc, le chapitre 5 concerne l'application de la métabolomique non ciblée à des domaines peu abordés jusqu'à présent avec une vision interdisciplinaire, comme la découverte de composés chimiques utilisés dans la communication entre individus d'une même espèce. Plus précisément, l'objectif de ce chapitre était la détermination des composés émis par les souriceaux qui induisent une activation de l'instinct maternel. La principale nouveauté qui devait être réalisée portait sur la phase d'échantillonnage, car non seulement il fallait travailler avec des animaux vivants, mais les conditions d'échantillonnage ne devaient pas interférer avec le développement des souriceaux ou induire un stress excessif pouvant affecter l'étude. D'une part, l'**Article scientifique V** se concentre sur les composés de nature plus volatile, adaptant l'extraction DHS-P&T pour capturer du volatolome d'animaux vivants, qui a ensuite été analysé par GC-MS. L'identité de 11 composés volatils provenant de souriceaux potentiellement impliqués dans l'induction des soins maternels chez les

femelles adultes a été déterminée. D'autre part, l'**Article scientifique VI** se concentrait sur les composés de nature moins volatile. Des échantillons ont été prélevés en frottant doucement des écouvillons sur les zones faciales et anogénitales des souriceaux. Les extraits ont été analysés par LC-IMS-HRMS et 5 composés ont été provisoirement identifiés à partir de souris néonatales avec une possible implication dans la communication mère-enfant.

CONTENTS

OBJECTIVES..... 1

CHAPTER 1. GENERAL INTRODUCTION..... 9

1.1 Introduction to metabolomics..... 13

1.2 Scientific article I..... 17

1.3 References..... 73

CHAPTER 2. FOOD PROCESSING..... 75

2.1 Introduction..... 79

2.2 Scientific article II..... 81

2.3 Discussion of the results..... 109

2.4 References..... 113

CHAPTER 3. FOOD INTAKE..... 115

3.1 Introduction..... 119

3.2 Scientific article III..... 121

3.3 Medium-term biomarkers of orange intake..... 159

3.4 References..... 165

CHAPTER 4. FOOD AND HEALTH EFFECTS..... 169

4.1 Introduction..... 173

4.2 Scientific article IV..... 175

4.3 Discussion of the results..... 223

4.4 References..... 227

CHAPTER 5. INTRA-SPECIE CHEMICAL COMMUNICATION..... 229

5.1 Introduction..... 233

5.2 Scientific article V..... 237

5.3 Scientific article VI..... 269

5.4 Discussion of the results..... 305

5.5 References..... 309

CHAPTER 6. CONCLUSIONS AND FUTURE WORK..... 313

ANNEX I. CO-AUTHOR AGREEMENT..... 333

LIST OF ACRONYMS

$[M+H]^+$	Protonated molecule
$[M-H]^-$	Deprotonated molecule
ACN	Acetonitrile
BFI	Biomarkers of food intake
CE	Collision energy
CCS	Collision cross section ($Å^2$)
Da	Dalton
DDA	Data dependent acquisition
dd-MS^2	Data-dependent-MS-two, DDA acquisition mode for Thermo Sci. Q-OT instruments
DHS-P&T	Dynamic headspace with sorbent trapping
DIA	Data independent acquisition
DT	Drift time
DTIMS	Drift tube ion mobility instruments
EI	Electron ionization
EIC	Extracted ion chromatogram
ESI	Electrospray ionization
FWHM	Resolution at full width at half maximum, used for HRMS instruments
GC	Gas chromatography
GC-MS	Gas chromatography coupled to mass spectrometry
HDMS^E	High definition MS^E
HE	High energy function in the MS^E acquisition mode for Waters Corp. Q-TOF instruments

HILIC	Hydrophilic interaction liquid chromatography
HPLC	High-performance liquid chromatography
HRMS	High resolution mass spectrometry
IMS	Ion mobility spectrometry/separation
IMS-HRMS	High resolution mass spectrometry with ion mobility separation
LC	Liquid chromatography
LC-MS	Liquid chromatography coupled to mass spectrometry
LE	Low energy function in the MS^E acquisition mode for Waters Corp. Q-TOF instruments
m/z	Mass to charge ratio
MeOH	Methanol
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MS^E	MS-to-the-E, DIA mode for Waters Corp. QTOF instruments
NAFLD	Non-alcoholic fatty liver disease
NPLC	Normal phase liquid chromatography
OPLS-DA	Orthogonal PLS-DA
OT	Orbitrap mass analyser
p(corr)	correlation predictive loading values in S-plot
p[1]	covariance loading values in S-plot
PCA	Principal component analysis
PLS-DA	Partial least square – discriminant analysis
ppb	Part per billion
ppm	Part per million
ppt	Part per trillion

Q	Quadrupole mass analyser
Q-HRMS	Hybrid quadrupole-high resolution mass spectrometry
Q-OT	Hybrid quadrupole-Orbitrap mass analyser
Q-TOF	Hybrid quadrupole-time of flight mass analyser
QC	Quality control, in metabolomics QC samples
RPLC	Reversed phase liquid chromatography
RT	Retention time
S/N	Signal to noise ratio
S-plot	model visualization derived from OPLS-DA
TOF	Time of flight mass analyser
TWIMS	Travelling wave ion mobility instruments
UHPLC	Ultra-high-performance liquid chromatography

Objectives

Objectives

The **main objective** of this doctoral thesis is the exploration and implementation of untargeted metabolomic strategies in the fields of food and health, together with advanced analytical techniques based on the coupling of chromatographic separations, both liquid (LC) and gas (GC) chromatography, with low and high resolution mass spectrometry (MS). Therewith, it is intended to make progress on the knowledge that this approach contributes to these highly relevant research fields.

Based on this general objective, some **specific objectives** have been established for each of the works included in the present thesis:

- ◇ Application of untargeted metabolomics to three different areas of food: food authentication, food intake and food effects on health.
- ◇ Obtaining a classification model for smoked fish products by means of the volatile profile obtained by dynamic extraction using purge-and-trap (DHS-P&T) and analysed by GC-MS.
- ◇ Determination of biomarkers of orange intake in blood plasma analysed by LC coupled to ion mobility separation (IMS) in combination with high resolution mass spectrometry (HRMS).
- ◇ Determination of metabolic changes in rat liver, with induced non-alcoholic fatty liver disease (NAFLD), by the simultaneous intake of compounds with reported antioxidation effect (resveratrol and pterostilbene).
- ◇ Determination of biomarkers related to the chemical signalling and behaviour associated with the maternal response in mice, emitted by their pups in the first days of age.
- ◇ Study the potential of IMS in untargeted metabolomics and in the biomarkers identification.

It is noteworthy that this work has been carried out thanks to close collaborations with research groups or companies from different fields. These collaborations and/or interactions have been essential for the development

of the thesis and have made clear the importance of multidisciplinary collaboration in scientific research, and the wealth of knowledge it contributes. Chapter 2 has been carried out thanks to the collaboration with the company Sea Delight Europe, S.L, with special mention to Mª Carmen Corell and Fernando Domínguez. Chapter 3 was developed thanks to the collaboration of the Department of Preventive Medicine and Public Health, Food Sciences, Toxicology and Legal Medicine of the University of Valencia, led by Prof. Dr. Dolores Corella. Chapter 4 arose from the collaboration with Prof. Dr. María Puy Portillo and Dr. Alfredo Fernández from the. Last but not least, the studies compiled in the fifth chapter were completed thanks to the close collaboration with the Laboratory of Functional Neuroanatomy from the Predepartmental Unit of Medicine from *Universitat Jaume I*, led by Prof Dr. Fernando Martínez García. In addition, during the period of doctoral studies, an international stay was carried out at the *Service de Pharmacologie et d'Immunoanalyse, CEA Paris-Saclay* (France), a renowned centre in the field of metabolism and metabolomics, where advanced analytical methodologies have been developed in HRMS directed to metabolomics studies.

Objetivos

El **objetivo principal** de la presente tesis doctoral es la exploración e implementación de las estrategias de la metabolómica no dirigida en el campo de la alimentación y la salud mediante el uso de en conjunto con técnicas analíticas avanzadas basadas en el acoplamiento de separaciones cromatográficas, tanto cromatografía líquida (LC) como de gases (GC), con la espectrometría de masas (MS) tanto de alta como de baja resolución. Con ello, se pretende avanzar en el conocimiento que aporta esta aproximación a estos campos de investigación de gran relevancia.

Basándose en este objetivo general, se han establecido una serie de **objetivos específicos** para cada uno de los trabajos incluidos en la presente tesis:

- ◇ Aplicación de la metabolómica no dirigida a tres ámbitos distintos relacionados con la alimentación: autentificación de alimentos, ingesta y efectos de la alimentación en la salud.
- ◇ Obtención de un modelo de clasificación de productos pesqueros ahumados mediante el perfil de volátiles obtenidos a través de extracción dinámica por purga y trampa (DHS-P&T) y analizados por GC-MS.
- ◇ Determinación de biomarcadores de ingesta de naranja en plasma sanguíneo por medio de análisis por LC acoplada a separación por movilidad iónica (IMS) en combinación con espectrometría de masa de alta resolución (HRMS).
- ◇ Determinación de los cambios a nivel metabólico en hígados de ratas, con enfermedad del hígado graso no alcohólico (NAFLD) inducida, por la ingesta simultanea de compuestos con efecto antioxidante reportado (resveratrol y pterostilbeno).
- ◇ Determinación de biomarcadores relacionados con la señalización química y conducta asociada con la respuesta maternal de ratones emitidos por las crías en los primeros días de edad.

- ◊ Estudiar el potencial de la IMS en la metabolómica no dirigida y en la identificación de biomarcadores.

Cabe destacar, que este trabajo se ha llevado a cabo gracias a estrechas colaboraciones con grupos de investigación o empresas de distintos ámbitos. Dichas colaboraciones y/o interacciones han resultado esenciales para el desarrollo de la tesis y han dejado patente la importancia de la colaboración multidisciplinar en la investigación científica, y la riqueza de conocimiento que aporta. El capítulo 2 se ha podido llevar a cabo gracias a la colaboración con la empresa Sea Delight Europe, S.L, con mención especial a Mª Carmen Corell y a Fernando Domínguez. El capítulo 3 se desarrolló gracias a la colaboración el Departamento de Medicina Preventiva y Salud Pública, Ciencias de la Alimentación, Toxicología y Medicina Legal de la Universidad de Valencia dirigido por la catedrática Dolores Corella. El capítulo 4 surgió de la colaboración con la catedrática María Puy Portillo y el Dr. Alfredo Fernández del Departamento de Nutrición y Ciencia de los Alimentos de la universidad del País Vasco. Por ultimo y no menos importante, los estudios recopilados en el quinto capítulo se llevaron a término gracias a la estrecha colaboración con el Laboratorio de Neuroanatomía Funcional dirigido por el catedrático de la Unidad Predepartamental de Medicina de la Universidad Jaume I Fernando Martínez García. Además, durante el periodo de estudios de doctorado se realizó una estancia internacional en el *Service de Pharmacologie et d'Immunoanalyse, CEA Paris-Saclay* (Francia), centro de renombre en el ámbito de metabolismo y metabolómica, donde se desarrollaron metodologías analíticas avanzadas en HRMS dirigidas a estudios de metabolómica.

Objectifs

L'**objectif principal** de cette thèse doctorale est l'exploration et l'implémentation des stratégies de métabolomique non ciblée dans le domaine de l'alimentation et de la santé, grâce à l'utilisation combinée de techniques analytiques avancées basées sur le couplage de séparations chromatographiques, à la fois la chromatographie liquide (LC) et gazeuse (GC), avec la spectrométrie de masse (MS) haute et basse résolution. Avec cela, visant à faire progresser la connaissance que cette approche apporte à ces domaines de recherche de grande pertinence.

Sur la base de cet objectif général, une série d'**objectifs spécifiques** ont été établis pour chacun des travaux inclus dans cette thèse :

- ◇ Application de la métabolomique non ciblée à trois domaines différents de l'alimentation : authentification des aliments, consommation et effets des aliments sur la santé.
- ◇ Obtention d'un modèle de classification des produits de la pêche fumé par le biais du profil de composants volatils obtenu par extraction dynamique de l'espace de tête avec piégeage par sorbant (DHS-P&T) et analysé par GC-MS.
- ◇ Détermination de biomarqueurs de consommation d'orange dans le plasma sanguin par analyse LC couplée à la séparation par mobilité ionique (IMS) en combinaison avec la spectrométrie de masse haute résolution (HRMS).
- ◇ Détermination des changements au niveau métabolique dans le foie de rats, atteints de stéatose hépatique non alcoolique induite (NAFLD), par la prise simultanée de composés ayant un effet antioxydant rapporté (resvératrol et ptérostilbène).
- ◇ Détermination de biomarqueurs liés à la signalisation chimique et au comportement associés à la réponse maternelle des souris émis par leur progéniture dans les premiers jours de l'âge.
- ◇ Etudier le potentiel de l'IMS en métabolomique non ciblée et dans l'identification de biomarqueurs.

Il est à noter que ce travail a été réalisé en collaborations étroites avec des groupes de recherche ou des entreprises de différents domaines. Ces collaborations et/ou interactions ont été essentielles pour le développement de cette thèse et ont mis en évidence l'importance de la collaboration multidisciplinaire dans la recherche scientifique, et la richesse des connaissances que ceci apporte. Le chapitre 2 a été réalisé grâce à la collaboration de la société Sea Delight Europe, S.L, avec une mention spéciale à Mª Carmen Corell et Fernando Domínguez. Le chapitre 3 a été développé grâce à la collaboration du Département de médecine préventive et de santé publique, des sciences alimentaires, de toxicologie et de médecine légale de l'Université de Valence, dirigé par le Professeur Dolores Corella. Le chapitre 4 est né de la collaboration avec le Professeur María Puy Portillo et le Docteur Alfredo Fernández du Département de nutrition et des sciences alimentaires de l'Université du Pays Basque. Finalement, les études compilées dans le cinquième chapitre ont été réalisées grâce à l'étroite collaboration avec le Laboratoire de neuroanatomie fonctionnelle de l'Unité prédépartementale de médecine de l'Université Jaume I dirigé par le Professeur Fernando Martínez García. En outre, pendant la période des études doctorales, un séjour international a été effectué au Service de Pharmacologie et d'Immunoanalyse, CEA Paris-Saclay (France), un centre renommé dans le domaine du métabolisme et de la métabolomique, où des méthodes analytiques avancées ont été développées avec HRMS dirigé aux études de métabolomique.

Chapter 1.
General introduction

Chapter 1. General introduction

1.1 Introduction to metabolomics

1.2 Scientific article I

"Chromatography hyphenated to high resolution mass spectrometry in untargeted metabolomics for investigation of food (bio)markers"

Trends in Analytical Chemistry 135 (2021) 116161

1.3 References

1.1 Introduction to metabolomics

What is Metabolomics all about?

There is a growing interest in increase the knowledge about how an ongoing biological process may affect an organism (biological functionality) and the molecules involved on the biological mechanisms (biological components), attending to the System Biology paradigm. This is a holistic and interdisciplinary approach based on the study of the complex interactions that occur withing a biological system (cells, organs, organisms or even groups of organisms) (Veenstra, 2021).

In this field of study, *-omics* strategies have caused huge progress. -O*mics* strategies might be divided in genomics, transcriptomics, proteomics and metabolomics (**Figure 1**); and their primary aim is the identification of a set of biomolecules (genes, transcripts, proteins and metabolites, respectively) present in a specific biological sample and in a complete and integrated way (Weckwerth, 2003). From these "big four omics", a wave of other omic fields have been created focusing, for example, in a narrower set of biomolecules as epigenomics, glycomics, lipidomics or volatolomics; or by focusing in a determinate domain as microbiomics or foodomics.

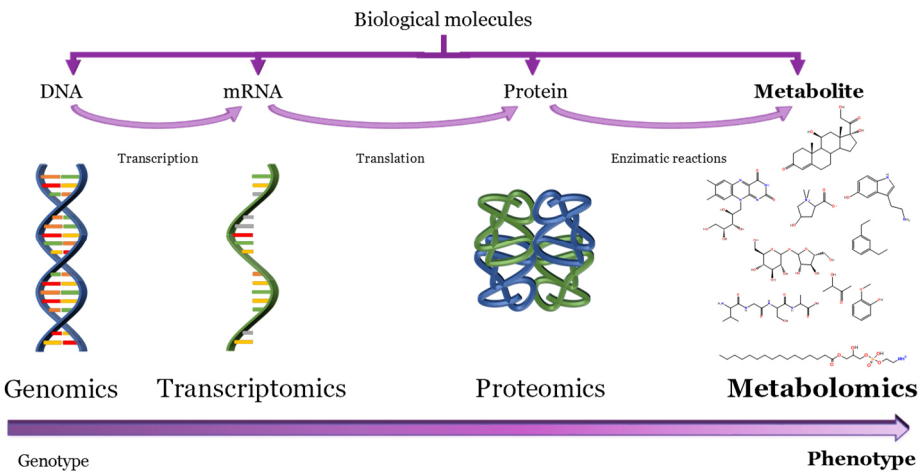


Figure 1. *Schematic representation of the main omics strategies.*

Within the context of metabolomics, this *-omic* strategy concerns the study of *metabolome* and its perturbation by the effect of endogenous and/or exogenous causes as disease, environment or dietary influences (Lindon et al., 2003). The term *metabolome* (Oliver et al., 1998) is used to address the whole set of low molecular weight molecules (≤ 1500 Da) metabolites, intermediates and metabolism products or even exogenous compounds present in a cell, tissue, organ or organism (Fiehn, 2002). Metabolites are in general the downstream products of the gene transcription and translation; and they reflect most closely the operation of a biological system, its phenotype. Therefore, metabolomics can provide an instantaneous and comprehensive snapshot of the metabolic state of the biological system under study.

Compared to the genomics, transcriptomics and proteomics methods that focus in a single chemical class of compounds, the analysis of metabolome presents a significant analytical challenge due to the disparate physical and chemical properties of the molecules that composes it. Moreover, their relative concentration may have a huge variation from a compound to another from 7 or 9 orders of magnitude of difference, exceeding sometimes the linear range of some analytical techniques employed (Cifuentes, 2013; Dunn & Ellis, 2005).

Due to this fact, metabolomics owes in large part its rise and expansion to the great technological development of analytical techniques and instrumentation. Nowadays, the most used analytical technique to analyse a large number of metabolites simultaneously is mass spectrometry coupled to different chromatographic separation techniques, such as liquid or gas chromatography, relegating to second place in this field the NMR mainly due to poorer sensitivity (Kuehnbaum & Britz-McKibbin, 2013). But, although nowadays the technology employed for the analysis is highly sensitive and sophisticated, due to the high diversity of chemical structures that compose the metabolome there is no a single technique to analyse the entire metabolome. Therefore, complementary approaches have to be employed in terms of extraction, separation, detection, quantification and identification to cover as many metabolites as possible (Bowen & Northen, 2010).

Another bottleneck in metabolomics is the extraction of the information and interpretation from the vast amount of data produced by the high-performance analysers. Sophisticated data treatment software and high-throughput chemometric tools (univariant and multivariant statistical analysis) are needed to unveil the relevant information (Karaman, 2017; Pinto, 2017).

Different metabolomic analytical approaches have been designed in order to answer different types of questions. On the one hand, *target metabolomics* or *metabolic profiling* is the approximation employed when the aim of the study is a pre-defined group of metabolites in a biological sample (e.g. a set of metabolites related to a specific pathway or a group of metabolites shared among different pathways) that are suspect to be involved in the analytical question of the study. In this approach the analytical methods are optimised for the identification and quantification of target compounds. On the other hand, *untargeted metabolomics* or *metabolic fingerprinting*, focuses in the differentiation and/or classification of samples based in the comparison of patterns or fingerprints of metabolites that change in response to an endogenous or exogenous factors, with the final aim of identify relevant biomarkers of the altered state under study, without the bias concerning the choice of targets to be analysed (Dettmer et al., 2007; Weckwerth, 2003).

This thesis mainly focuses on the application of untargeted metabolomics approaches involving mass spectrometry combined with liquid and gas chromatography. The **Scientific article I** hereunder explains in detail all the steps that are comprise in the untargeted metabolomics analytical workflow based on chromatography separation hyphenated to high resolution mass spectrometry. This article was specifically aimed at the discovery of food (bio)markers, and therefore shows the state of the art of this field and the most widely adopted strategies. A great part of the thesis has been carried out in this area, however this description of the untargeted metabolomic workflow and the different aspects to take into account is transversal to other most of the areas of study, such as health or chemical signalling and behaviour, which is the main theme of **Chapter 5**.

1.2 Scientific article I

Trends in Analytical Chemistry 135 (2021) 116161

Contents lists available at ScienceDirect

Trends in Analytical Chemistry

journal homepage: www.elsevier.com/locate/trac

Chromatography hyphenated to high resolution mass spectrometry in untargeted metabolomics for investigation of food (bio)markers

Leticia Lacalle-Bergeron, David Izquierdo-Sandoval, Juan V. Sancho, Francisco J. López, Félix Hernández, Tania Portolés*

Enviromental and Public Health Analytical Chemistry, Research Institute for Pesticides and Water (IUPA), University Jaume I, Av. Sos Baynat S/N, 12071 Castellón de la Plana, Spain

ARTICLE INFO

Article history:
Available online 24 December 2020

Keywords:
Untargeted metabolomics
Mass spectrometry
LC-HRMS
GC-HRMS
Food sciences
Nutrition
Biomarkers

ABSTRACT

Currently, there is a growing demand by our society, authorities and science to increase the knowledge about the quality of food and its relationship with health and disease. Untargeted metabolomics approaches are emerging as powerful tools for exploring metabolic changes in biological systems under different conditions with great potential in the food field. To this aim, it is necessary to apply advanced analytical techniques, such as chromatography hyphenated to high resolution mass spectrometry, which provides enough sensitivity and selectivity to cover a wide range of metabolites in complex samples, as food and biological samples. The objective of this work is to provide an overview of the most widely adopted strategies based on the use of high resolution mass spectrometry-based techniques for the identification of food (bio)markers through the untargeted metabolomics workflow. Detailed information is provided about the trends in each stage of the metabolomics process from updated literature with the objective to help researchers to select the most appropriate metabolic approaches.

1. Introduction

Foodomics has been defined as a new discipline that studies food and nutrition domains combining the application of advanced analytical techniques (omics tools) and bioinformatics. The use of omics tools, such as genomics, transcriptomics, proteomics and/or metabolomics, is a requirement to address the challenges presented in emerging working areas included in foodomics studies [1]. Metabolomics can be defined as a non-selective, comprehensive analytical approach for the identification and quantification of metabolites in a biological system, typically those small molecules with a molecular weight below 1500 Da [2]. Metabolomics has become a powerful tool for the study of the complex interactions between diet and the human or animal organisms enabling to expand our knowledge of the subtle changes at metabolic level activated by foods, nutrients and disease. It has allowed significant improvements in the field of dietary assessment since it enables the identification of novel and robust biomarkers of food intake (BFIs) enhancing the accuracy and the objectivity in the measurement of dietary exposures and reducing the bias and errors associated with self-report methods [3]. On the other hand, the potential of metabolomics as a robust, efficient and sensitive analytical methodology in food safety, quality and traceability is widely recognized [4].

Metabolomic studies are challenging because of the aim to characterize complex and diverse biological matrices containing compounds with a wide range of polarities or volatilities. Carbohydrates, lipids, amino acids, amines, steroids, phenolic compounds, carotenoids, alkaloids or volatile compounds, are examples of compounds that constitute the metabolome [2]. This enormous diversity has led to the emergence of sub-areas within the metabolomics field to narrow down the search for compounds with similar physicochemical properties. As an example, lipidomics deals with the determination of lipid classes, subclasses and lipid signalling molecules, providing a tool for the assessment of changes in lipid metabolism [5]. On the other hand, volatolomics is the subunit of metabolomics responsible of the detection, characterization and quantification of volatile metabolites in a biological system [6].

In general, two complementary approaches are used in metabolic research: metabolic profiling (targeted metabolomics) and metabolic fingerprinting (untargeted metabolomics). Metabolite profiling focuses on the analysis of a group of metabolites such as

* Corresponding author.
E-mail address: tportole@uji.es (T. Portolés).

https://doi.org/10.1016/j.trac.2020.116161
0165-9936/

Trends in Analytical Chemistry 135 (2021) 116161

Chromatography hyphenated to high resolution mass spectrometry in untargeted metabolomics for investigation of food (bio)markers

Leticia Lacalle-Bergeron, David Izquierdo-Sandoval, Juan V. Sancho, Francisco J. López, Félix Hernández, Tania Portolés*

Enviromental and Public Health Analytical Chemistry, Research Institute for Pesticides and Water (IUPA), University Jaume I, Av. Sos Baynat S/N, 12071 Castellón de la Plana, Spain

Highlights

- Updated review (2017–2020) of untargeted metabolomics strategies applied on food field.
- Advances in LC, GC and HRMS greatly contributed to food (bio)markers research.
- Trends and future of HRMS-based untargeted metabolomics strategies are discussed.
- Applications on food processing, food intake and food health effects are reviewed.

Abstract

Currently, there is a growing demand by our society, authorities and science to increase the knowledge about the quality of food and its relationship with health and disease. Untargeted metabolomics approaches are emerging as powerful tools for exploring metabolic changes in biological systems under different conditions with great potential in the food field. To this aim, it is necessary to apply advanced analytical techniques, such as chromatography hyphenated to high resolution mass spectrometry, which provides enough sensitivity and selectivity to cover a wide range of metabolites in complex samples, as food and biological samples. The

objective of this work is to provide an overview of the most widely adopted strategies based on the use of high resolution mass spectrometry-based techniques for the identification of food (bio)markers through the untargeted metabolomics workflow. Detailed information is provided about the trends in each stage of the metabolomics process from updated literature with the objective to help researchers to select the most appropriate metabolic approaches.

Keywords

Untargeted metabolomics; Mass spectrometry; LC-HRMS; GC-HRMS; Food sciences; Nutrition; Biomarkers

Abbreviations

APCI	atmospheric pressure chemical ionization	LC	liquid chromatography
BFI	biomarker of food intake	LR	low resolution
CI	chemical ionization	MS	mass spectrometry
DDA	data dependent acquisition	MS/MS	tandem mass spectrometry
DIA	data independent acquisition	MS^n	sequential mass spectrometry
EI	electron ionization	NMR	nuclear magnetic resonance
ESI	electrospray ionization	OT	Orbitrap mass analyser
FS	full scan	Q	quadrupole mass analyser
GC	gas chromatography	RPLC	reversed phase liquid chromatography
HILIC	hydrophilic interaction chromatography	TOF	time-of-flight mass analyser
HRMS	high-resolution mass spectrometry	xC-HRMS	chromatographic techniques coupled to high resolution mass spectrometry
IMS	ion mobility spectrometry		
IT	ion trap mass analyser		

1. Introduction

Foodomics has been defined as a new discipline that studies food and nutrition domains combining the application of advanced analytical techniques (omics tools) and bioinformatics. The use of omics tools, such as genomics, transcriptomics, proteomics and/or metabolomics, is a requirement to address the challenges presented in emerging working areas included in foodomics studies [1]. Metabolomics can be defined as a non-selective, comprehensive analytical approach for the identification and quantification of metabolites in a biological system, typically those small molecules with a molecular weight below 1500 Da [2]. Metabolomics has become a powerful tool for the study of the complex interactions between diet and the human or animal organisms enabling to expand our knowledge of the subtle changes at metabolic level activated by foods, nutrients and disease. It has allowed significant improvements in the field of dietary assessment since it enables the identification of novel and robust biomarkers of food intake (BFIs) enhancing the accuracy and the objectivity in the measurement of dietary exposures and reducing the bias and errors associated with self-report methods [3]. On the other hand, the potential of metabolomics as a robust, efficient and sensitive analytical methodology in food safety, quality and traceability is widely recognised [4].

Metabolomic studies are challenging because of the aim to characterize complex and diverse biological matrices containing compounds with a wide range of polarities or volatilities. Carbohydrates, lipids, amino acids, amines, steroids, phenolic compounds, carotenoids, alkaloids or volatile compounds, are examples of compounds that constitute the metabolome [2]. This enormous diversity has led to the emergence of sub-areas within the metabolomics field to narrow down the search for compounds with similar physicochemical properties. As an example, lipidomics deals with the determination of lipid classes, subclasses and lipid signalling molecules, providing a tool for the assessment of changes in lipid metabolism [5]. On the other hand, volatolomics is the sub-unit of metabolomics responsible of the detection, characterization and quantification of volatile metabolites in a biological system [6]. In general, two complementary approaches are used in metabolic research: metabolic profiling (targeted metabolomics) and

metabolic fingerprinting (untargeted metabolomics). Metabolite profiling focuses on the analysis of a group of metabolites such as those related to a specific metabolic pathway. In this approach, target metabolites are selected beforehand and they are assessed using specific analytical methods. Technological advances have increased the number of metabolites that can be quantified simultaneously. Moreover, the results of metabolic profiling are independent of the technology used for data acquisition. Metabolic fingerprinting does not aim to identify the entire set of metabolites but rather to compare patterns or fingerprints of metabolites that change in response to an altered state promoted by endogenous (disease, genetics…) or exogenous (diet, environment…) conditions. It can be used as a tool to evaluate the state of a biological system by comparing, for example, control and disease subjects, or to assay the success of a particular treatment (prognosis/recovery). Once a differential pattern is discovered, further steps to identify the contributing compounds (qualitative) and to determine the absolute amounts of metabolites that participate in the processes studied (quantitative) must be followed. This issue is not trivial and prior to boarding on the task of discovering metabolic biomarkers, sufficiently sensitive and selective instruments and extensive compound libraries for metabolite identification must be available, while wide experience in data analysis and interpretation are also necessary [7]. Unlike the traditional analytical workflow, untargeted metabolomics is an hypothesis-driven methodology, that means that to address a biological question the experiment must be design with the broadest perspective as possible, and the hypothesis is generated from the result [5]. As large data sets are obtained from the results, potent statistical tools, as multivariate analysis, are necessary to reduce the data complexity and to reveal underlying trends from which it is hoped that hypothesis can be generated [8]. **Fig. 1** shows a typical workflow followed in metabolomics fingerprinting. Concerning the detection and identification of metabolites, high resolution mass spectrometry (HRMS)-based techniques are, undoubtedly, the most suitable option to deal with the vast diversity of small molecules with distinct physicochemical properties in complex biological matrixes that constitute the metabolome. The main advantages of HRMS-based metabolomics are the high sensitivity and selectivity as well as the accurate-mass full-spectrum acquired data, together with possibility to be

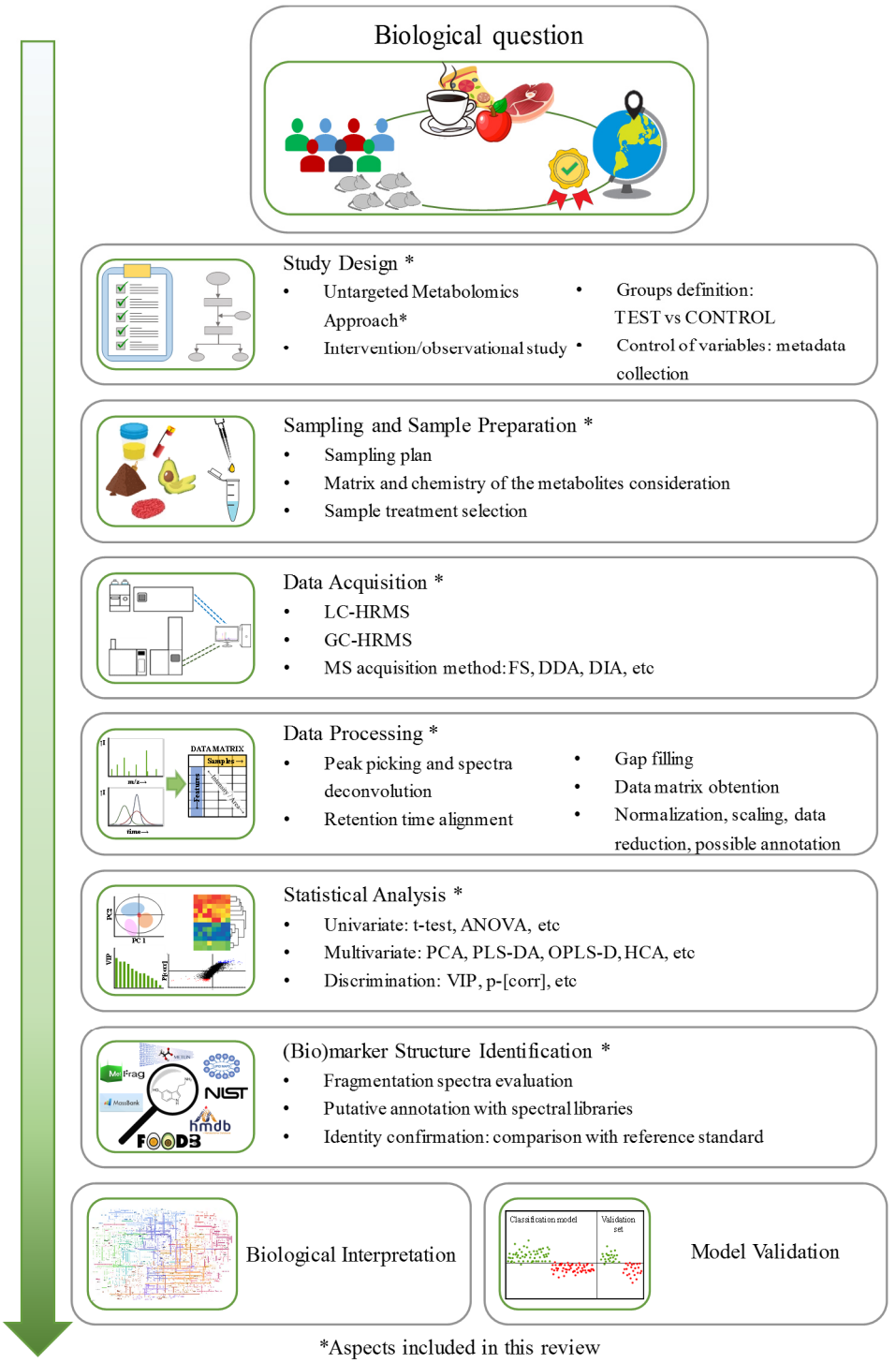


Fig. 1. *General overview and schematic content of the untargeted metabolomics workflow based on xC-HRMS analysis.*

coupled on-line to a separation technique. The hyphenation of separation techniques, mainly gas chromatography (GC) and liquid chromatography (LC), with HRMS reduces the complexity of the mass spectral data, enhancing the sensitivity of the detection and providing additional information about the physicochemical characteristics of the analysed molecules. Moreover, HRMS analysers can be used as a hybrid instrument allowing acquisitions in tandem mass spectrometry mode (MS/MS or MS^n) incorporating fragmentation data of the metabolites and facilitating the confirmation of known, reported, compounds or assisting the elucidation of unknown metabolites.

The starting point of this work was a comprehensive search in Scopus database using the following keywords: “precision foods”, “functional foods”, “precision nutrition”, “food intake”, “biomarkers of intake”, “nutritional assessment”, “dietary markers”, “nutrimetabolomics”, “food quality”, “food safety”, “food authenticity”, “food fraud” and “food traceability”; along with keywords related to the analytical technique and the methodology: “untargeted” (and synonyms “fingerprinting”, “untarget” and “non-target”), “metabolomics”, “mass spectrometry”, “HRMS” and “MS”; in papers published between 2017 and 2020. Reviews, trends, perspectives and book chapters were kept separately as a source of information. With the articles selected, a discussion of the trends in chromatography-HRMS-based metabolomics fingerprinting within in the context of foodomics is provided using as guideline the workflow shown in **Fig. 1**.

2. Study design

Bearing in mind the objective of metabolomics fingerprinting, the experimental design requires careful consideration prior to laboratory work to ensure the quality and validity of the results. Within this approach, an appropriate experimental design must undertake the acquisition of data related to a specific biological question while ensuring that covariates or cofounders are not present or are well characterised [9].

A common feature in experimental design is that cohorts should be homogenous in those factors that are not included in the biological question, avoiding unnecessary errors, false leads and “statistical noise” [8]. For

example, within the context of the discovery of fraudulent practises in the food industry, dead on arrival and regularly slaughtered chickens metabolite patterns were compared. All the chickens were grown on the same farm, were of the same age and were fed the same diet; moreover, the same tissues were analysed [10]. Sometimes, sample characterisation after sample collection is necessary to define in which cohort the sample belongs; for example, testing panels made up of professional tasters were used to assess quality in olive oil samples [11] and green tea samples [12]. In other study, the quality in berries of sea buckthorn was defined using a colorimeter [13]. In some occasions, when the objective is to look for the variability between geographic origin or variety, the characterisation is provided by the supplier [14].

A common feature in experimental design is that cohorts should be homogenous in those factors that are not included in the biological question, avoiding unnecessary errors, false leads and "statistical noise" [8]. For example, within the context of the discovery of fraudulent practises in the food industry, dead on arrival and regularly slaughtered chickens metabolite patterns were compared. All the chickens were grown on the same farm, were of the same age and were fed the same diet; moreover, the same tissues were analysed [10]. Sometimes, sample characterisation after sample collection is necessary to define in which cohort the sample belongs; for example, testing panels made up of professional tasters were used to assess quality in olive oil samples [11] and green tea samples [12]. In other study, the quality in berries of sea buckthorn was defined using a colorimeter [13]. In some occasions, when the objective is to look for the variability between geographic origin or variety, the characterisation is provided by the supplier [14].

Regardless the aim of the study, the collection of metadata during sample collection is crucial to avoid bias and the incorporation of data related to confounding factors into the statistical analysis [9]. In studies involving humans it is imperative to collect demographic/physiological/lifestyle metadata since such factors are difficult to control and inter-subject metabolic variation could be hiding the metabolic changes produced by the food or diet to be assessed. In animal or cellular assays, where there is commonly an extensive control over these factors, the inter-subject variation is usually negligible. Inclusion and exclusion criteria are commonly

established in human studies to avoid incorporating subjects whose demographic/physiological/lifestyle characteristics can produce undesirable results in a specific study. For example, smokers used to be excluded due to the potential exposure to polycyclic aromatic hydrocarbons [15].

Two approaches can be considered when designing foodomics studies: intervention studies and observational studies. Dietary intervention study designs generally involve participants consuming a specific standardised diet or food product over a defined time. In this way, the variation introduced by food storage, preparation process, as well as the type of food and the nutritional value, is in usually controlled. Biofluids, urine or blood, are collected at specific time-points depending on the research interest. For example, blood samples were collected at baseline and after three and six weeks of treatment with the aim to compare the metabolite fingerprints at different levels of red-meat consumption [15]. On the other hand, twenty-four-hour urine samples were collected in a four-way cross-over intervention for the investigation of biomarkers in different kinds of meat consumed in a restricted diet during 48 h [16]. Within a cross-over design, such as the previous example, the participants receive all treatments reducing the inter-subject variation. When a succession of treatments is applied to the same participants is necessary to include washouts in the study design, which can consist in returning to the habitual diet or in excluding the food of interest for the study. The importance of washouts lies in returning to the basal metabolic levels avoiding carryover. Besides, in such studies that include blood collection, the washout duration must be longer to ensure the recovery of red blood cells and platelets [17]. Other elements to highlight are the randomisation of subjects and the nutritional and isocaloric equivalence between treatments. Generally, dietary intervention studies are expensive and laborious to conduct, and some methodological compromises are required, such as limiting the sample size or reducing the time of study [3].

In observational studies, two groups, generally low and high consumers of the food(s) or diet of interest, are selected from food intake data collected by traditional dietary assessment methods such as food frequency questionnaires (FFQ), dietary diaries or other dietary assessment tools. Broadly, participants are selected from large cohorts to perform a cross-

sectional study; in other words, groups of participants are compared at a single time point. For example, a cross-sectional design was applied to a subgroup of the SU.VI.MAX cohort, funded by the French National Cancer Institute. All participants were invited to complete a 24-h dietary record every two months up to a total of 10, covering all days of the week and all seasons of the year to assess their adherence to the French dietary recommendation [18]. A meticulous exclusion process was applied in the previous example, for selecting a limited number of participants from a large cohort. A stratified randomisation was performed to ensure that experimental groups are balanced concerning the confounding variables [19]. Observational work usually involves studies with a large number of samples and long study time; however, the limitations of traditional dietary assessment in providing reliable information could be a source of bias. Sometimes, cross-sectional studies in large cohorts are used to validate biomarkers identified by interventional designs [20].

3. Sampling and sample preparation

Once the experimental design is established, the next steps in the analytical process involve sample collection and sample preparation in the laboratory, including the shipping and storage of samples. It is essential to minimize sources of confounding factors, random or systematic errors during these stages to ensure the generation of robust and reproducible data, which only result in the variation between the different classes defined in the study design. Samples should be representative in terms of the biological question, defining factors such as the type and amount of sample, time of collection, and ensuring proper randomisation and group balancing within the sampling plan [21]. After collection of the sample, the metabolome may change because of many factors such as enzymatic activities, exposure to oxygen, UV light and temperature; so, optimum transport and storage conditions must be established to avoid sample losses, transformations, or contamination. Sample preparation in untargeted metabolomics aims transforming the physiochemical properties of the sample in a reproducible way to make it compatible with the analytical method. It should be as less selective as possible, maintaining the most the metabolomic composition of the sample, covering a wide range of compounds. Minimizing the steps in sample

preparation avoids losses of metabolites and facilitates a high sample throughput [22]. In most cases, sample preparation is reduced to a straightforward solvent-extraction [4] or even a simple "dilute and shoot" in the case of less complex matrices when using LC separation [22].

Before the extraction, it is necessary to homogenise the sample and reduce its size, together with metabolism quenching. In solid samples (e.g. food, human and animal tissues), freeze-dried powder sample or frozen samples are commonly used; the sample can also be ground in a mortar with liquid nitrogen [23]. Vortex and ultrasounds sonication can be used to perform a more exhaustive extraction of metabolites during solvent-extraction in solid matrices [[24], [25], [26]], while in liquid samples (e.g. beverages and bio-fluids), stirring and aliquots are usually applied for homogenization and size reduction.

Hydro-organic mixtures containing water, methanol, acetonitrile and/or formic acid are a common choice for extraction since such versatile solvent systems provide enough solubility for covering polar and semipolar metabolites [21]. Hydrophobic extraction mixtures using organic solvents, such as chloroform or dichloromethane, are appropriate for the extraction of the non-polar fraction of the metabolome, as for example lipids or volatile compounds. Double extractions are sometimes applied to cover both polar and non-polar metabolites [10,27]. For example, chloroform, water and methanol was applied for freeze-dried carrot samples; after centrifugation, the aqueous phase (water/methanol) was used for analysing polar compounds, and the chloroform phase was evaporated and reconstituted with methanol for lipids analysis [28].

Cold extraction is recommended to avoid enzymatic activity. Some compounds can be used to stop the metabolomic activity, as long as global extraction is not compromised; for example, O-(carboxymethyl) hydroxylamine hemihydrochloride (OCMHA) was added to inhibit enzyme *alliinases* in garlic samples [14]. After extraction, centrifugation is applied to eliminate the solid residues and the proteins precipitated by the organic solvent.

In GC-based methods, the non-volatile compounds should be carefully removed or being chemically derivatised, which increases the complexity of the sample treatment adding time-consuming steps. A typical derivatisation consist on two-steps process: methoxymation for ketone functional groups protection with methoxyamine, followed by silylation with reagents as BSTFA [27], MSTFA [29] and MTBSTFA [19], to reduce the polarity of the molecule by reacting with the active hydrogen of polar functional groups(–OH, –COOH, $–NH_2$, –SH and $–PO_4^{-3}$). In this way, it was possible to detect simultaneously chemical families like amino acids, sugars, organic acids and some fatty acids, among other metabolites, in vegetable matrices [30] or human biofluids [31].

In volatolomics studies, GC analysis is the natural selection. Dynamic headspace purge and trap (DHS-P&T) has been used to extract the volatile fraction in olive oil samples. Volatiles were released from the sample by the use of a nitrogen stream and then retained on a reversed-phase sorbent cartridge [11]. In some cases, a solvent extraction from the trap cartridge is applied with a GC-suitable solvent such as n-hexane. Another alternative for volatile extraction is to establish an equilibrium between the vapour phase and the adsorbent in a closed space. Thus, headspace solid-phase microextraction (HS-SPME) has been satisfactorily used in seeds and whisky analysis [32,33]. The volume of sample, temperature, equilibrium time, the necessity of stirring, salting-out and the type of fibre are parameters commonly optimised in SPME analysis, and PDMS/DVB coating is suitable for volatile global screening [34]. Parallel to the SPME analysis, a more extensive range of compounds, including volatiles and semi-volatiles, could be extracted with ethyl acetate in whiskey samples [33].

Blood plasma, serum and urine are the common biofluids studied. The extraction of metabolites from urine is usually made by dilution with water and centrifugation followed by filtration for removal urine proteins or particulates. The dilution can be done before or after the centrifugation, and the degree of dilution uses to be in the range 1:1 to 1:3 (*v/v*) [35]. An attempt to normalise the dilution factor was carried out by the measuring of specific gravity by refractometry before the analysis [36]. Regarding serum and plasma, due to their high protein content, the sample preparation scheme

involves a simple protein precipitation followed by centrifugation and reconstitution [[37], [38], [39]]. It is also possible to extract exogenous metabolites by using acidified methanol [18].

4. Instrumental analysis

There is no universal analytical technique in untargeted metabolomics. The analysis of complex samples and the vast diversity of small molecules with diverse physico-chemical properties that constitute the metabolome entails the need to use a wide variety of analytical techniques. It is highly recommended to run more than one platform to enhance the compound coverage and to obtain comprehensive information. The two major analytical platforms to perform untargeted metabolomics are nuclear magnetic resonance (NMR) spectroscopy and HRMS-based techniques. NMR advantages are the robust structural elucidation capabilities, the non-destruction of the sample and the detection of non-ionizable compounds, among others [40]. However, it is not capable of reaching the sensitivity of HRMS-based techniques and is less suitable to be coupled on-line to separation techniques. The hyphenation of separation techniques with HRMS reduces the complexity of the mass spectral data, enhancing the sensitivity of the detection, providing useful information about the physicochemical characteristics of the analysed compounds. In untargeted metabolomics, GC and LC are the most used separation techniques and both can be easily coupled to HRMS. The accurate-mass full-spectrum information provided by HRMS is essential for the reliable identification of the compounds previously separated by chromatography.

From the 79 articles reviewed in this paper that perform untargeted metabolomics for food related biomarkers, 68 used LC-HRMS, 7 GC-HRMS and only 4 a combination of both techniques. In other cases, one of these platforms is complementary to other techniques as capillary electrophoresis-HRMS, NMR o GC-MS (nominal mass analysers). This review focuses on the combination of chromatography with HRMS.

4.1. High resolution mass spectrometry (HRMS)

The progress in untargeted metabolomics has been mainly driven by the improvements of the analytical techniques; the most important being MS technological innovations [41]. The improvements have been mainly focused on the increase of mass resolving power and sensitivity, as well as broadening the dynamic range and enhancing the acquisition rate [4,42]. Among the HRMS analysers, time-of-flight (TOF) and the Orbitrap (OT) are the most used, while Fourier transform ion cyclotron resonance (FT-ICR) is less applied due to its low acquisition rate, which makes difficult the coupling with fast chromatographic separations, as well as to its higher maintenance costs, making it a less affordable analyser.

HRMS can be also used as hybrid instruments combined with low resolution (LR) mass analysers, such as ion trap (IT) or quadrupole (Q), allowing to work not only under full scan (FS) mode but also under tandem mass mode (MS/MS or MS^n), improving the identification based on the fragmentation patterns. TOF and hybrid Q-TOF instruments are the most used in untargeted metabolomics applied to food related sciences (applied in 11 % and 68 % of the reviewed research articles, respectively); followed by hybrid OT analysers (13 % using Q-OT, and 8 % using IT-OT). These mass analysers can achieve mass accuracy below 2 ppm (with internal calibration). Mass resolution expressed as full width at half maximum (FWHM) can reach values up to 80,000 and 1,000,000 for TOF-based and last generation OT-based instruments, respectively. However, the resolving power is dependent on the duty cycle for OT-based instruments while Q-TOF analysers are able to acquire at a scan rate up to 100 Hz independently of the resolving power. This fact has made Q-TOFs better suited when the chromatographic peaks are narrow as in GC or fast LC separations [43]. Nevertheless, the innovation on OT instruments has allowed the recent introduction of GC-OT instruments into the market with an interesting potential in future untargeted metabolomics applications [44]. Hybrid MS analysers allow the simultaneous MS acquisition and MS/MS or MS^n in a single injection (i.e. FS and target MS^n analysis). Under these acquisition modes, one can obtain both (semi)-quantitative (from the FS) and structural (from the MS^n) information in a

single injection. Data dependent and data independent acquisitions modes can be applied in analysis (DDA and DIA, respectively).

Under DDA, the instrument automatically switches from FS to MS/MS or MS^n of the preselected ions detected in the FS spectrum. This preselection is intensity dependent along with other predefined parameters and may negatively affect the DDA coverage specially for low abundance features [45]. Licha et al. satisfactorily applied Q-OT under DDA for analysis of mice plasma samples after application of ketogenic diet to study the metabolic profile and its relationship with tumour growth inhibition [46]. In the DDA, MS^2 criteria specified that the five most abundant ions from every scan cycle were isolated in the Q in a window of 0.8 *m/z* and subsequently fragmented. Tovar et al. implemented a DDA acquisition method in a Q-TOF instrument to study the effect of multifunctional diet in human metabolism where only the 4 most abundant ions from every precursor scan cycle were selected for fragmentation [47]. Nevertheless, there was the need to perform additional target MS/MS measurements for those potential markers that failed to be included in the previous DDA method.

DIA systematically performs the fragmentation of all precursor ions along the full *m/z* range (also called all-ion fragmentation (AIF) or MS^E among other commercial names) or within a selection of sequential mass windows (like SONAR or sequential window acquisition for all theoretical spectra (SWATH)). Although DIA covers the DDA limitation for low abundance ions, the resulting MS/MS spectra is a composite of fragment ions generated from all precursor ions. Thus, it is required the aid of powerful algorithms to stablish the link between the precursor ion and the fragmentation pattern [45]. Hoyos Ossa et al. applied MS^E acquisition method for the origin discrimination of Colombian green coffee [48]. The fragmentation spectra obtained under MS^E were not enough informative to allow the identity of the markers. Therefore, additional target MS/MS analysis was made to confirm the structure of the compounds used in the model of discrimination by origin. More information about data acquisition in untargeted metabolomics can be found in the extensive review of Fenaille et al. [42].

It is worth noting the difficulties to optimize a methodology in untargeted metabolomics when the compounds that may be relevant are unknown, contrary to targeted approaches, such as profiling metabolomics, where they are known, and reference standards are commonly available. Therefore, the choice is usually based on the possibility of fragmenting the maximum number of ions as possible and thus being able to cover a wider range of compounds that could be potential markers. Guo et al. made a comparative study of FS, DDA and DIA (AIF mode) in LC-QTOF untargeted metabolomics with different LC separations using spiked human urine samples. The best results where for FS in terms of sensitivity and quantitative precision, higher quality of MS^2 spectra with DDA but better MS^2 spectral coverage with DIA [45].

4.2. Gas chromatography-high resolution mass spectrometry (GC-HRMS)

GC is ideal for the separation of thermally stable and volatile compounds (or volatile derivatives previous chemical derivatisation). Capillary columns from non-polar stationary phases as 100 % dimethylpolisiloxane [31], to polar as 100 % polyethylenglicol [33] may be used. One of the most applied in untargeted metabolomics is the non-polar stationary phase 5 % dimethyl – 95 % dimethylpolysiloxane [27,29,30,49,50] or similar [11,19,32]. GC-MS is well established in metabolomics [41] because of its advantages of high chromatographic resolution, sensitivity and separation reproducibility [51]. However, aqueous samples must be dried or subjected to solvent exchange before GC-MS analysis (which can entail volatile losses). As mentioned above, those compounds that are not naturally volatile must be carefully removed or being chemically derivatised.

Electron ionization (EI), a robust and hard-ionization technique, is the most commonly used in GC-based metabolomic studies [52] where useful spectral databases have been built over the years, such as NIST. The availability of these databases facilitates the rapid identification of the markers by mass spectral matching, which makes it the main attractiveness of the GC-EI-MS, especially compared to the LC-MS based metabolomics [51].

The ionization source has notable impact on the mass analyser selected. Indeed, the significant in-source fragmentation makes a hybrid analyser less useful, and so GC-EI is commonly coupled to a single mass analyser as TOF working in FS acquisition. As illustrative example, the plasma metabolic profiles associated with meat and seafood consumption in Asian population [19] were studied by LC-QTOF and GC-EI-QTOF analysis (previous derivatisation). While for highlighted markers from LC analysis, additional MS/MS acquisition were needed for structural elucidation, GC-EI-MS analysis was performed only in FS and markers were annotated by fragmentation spectra matching with NIST library, with final identity confirmation with reference standards. The use of GC-EI-MS with LR analysers (e.g. Q) under FS mode continues to be widespread, since the structural identification power of the fragmentation spectrum together with the libraries make the exact mass acquisition of HRMS less necessary, in addition to being clearly more economical and accessible instrument for most laboratories.

Chemical ionization (CI) is less applied compared to EI. CI is a soft-ionization technique able to preserve the precursor ion, commonly limited to targeted analysis, as it is strongly dependent on the reagent gas and pressure used for the ionization [53]. Stupak et al. performed additional target MS/MS analysis with positive CI where the precursor ion was not found for those potential markers of quality and authenticity of Scotch whiskey with excessive fragmentation in EI [33].

4.3. Liquid chromatography-high resolution mass spectrometry (LC-HRMS)

LC is the most employed separation technique especially for aqueous samples as biofluids or some food matrices. Furthermore, it does not usually require complex sample preparations and involves short run times compared to GC. Due to the diversity of stationary phases and the different mobile phases that can be used, versatility is one of the main advantages of this technique, allowing its applicability to the analysis from medium to highly polar, low volatility and/or thermolabile compounds. If the interest is to reach the maximum coverage, as in untargeted metabolomics, more than one separation mechanisms should be assayed.

Reversed phase LC (RPLC) and hydrophilic interaction chromatography (HILIC) are the most used stationary phases in untargeted metabolomics. Among RPLC, ideally suited for the analysis of semi-polar/nonpolar analytes, C18 stationary phase is the most commonly used due to its well-known behaviour, good robustness and its ability to cover a wide range of compounds. It has been applied in the identification of biomarkers of intake [54,55], designation/geographical of origin [56,57], and the study of the effect of functional food in health [58,59], among others. Two complementary C18 RPLC strategies have been used to assess the changes in plasma metabolome by the consumption of an herbal supplement, one more geared towards the lipidic metabolome (lipidomics) and the other to general metabolome (metabolomics) [60]. C18 columns modified with polar endcapping (e.g. as HSS T3 from Waters) are becoming more popular, as they are able to support highly aqueous mobile phases (even 100 % water) expanding their applicability to more polar compounds compared to the traditional C18. Kozlowska et al. detected with this stationary phase some nitrogenous bases as tryptophan metabolites, organic acid and phase II metabolites in urine, usually difficult retain in conventional C18 [61].

HILIC seems to be the choice for polar to highly polar compounds, but insufficiently charged for ion-exchange chromatography. The stationary phase is a highly hydrophilic, such as silica or chemically modified silica (as amide) and the mobile phase is an organic solvent containing a small amount of water (at least 5 %) [62]. HILIC separations were applied for the analysis of polar lipids, in different life stages, of one of the most consumed seaweed for sushi (*Porphyra dioica*) [63] as well as for the assessment of garlic authenticity, detecting polar metabolites as phospholipids and small peptides and amino acids [14]. This separation mode is more affected by the chromatographic conditions and matrix effects, and it is known to be less reproducible than RPLC regarding retention time. HILIC is commonly used simultaneously with RPLC to obtain a complementary information on those polar compounds that RPLC cannot separate. As example, this combination has been applied for discovery of consumption biomarkers [36,64] and to study the effects of different diets on health [65,66]. Pérez-Miguez et al. highlighted the advantages of combining HILIC with RPLC (and even

capillary electrophoresis) for the study of coffee roasting process showing a comparative of the metabolites identified by each strategy [67].

Electrospray ionization (ESI) is clearly the preferred approach in untargeted metabolomics based on LC-MS analysis. Indeed, all the LC-HRMS studies reviewed made use of ESI, and 77 % of them used both positive and negative ionization modes. In comparison with GC-EI-MS, LC-ESI-MS is more affected by the instrument-to-instrument differences which makes troublesome the matching with mass spectral databases. This fact and the high quantity of non-reported compounds in LC databases may hamper the identification of the (bio)markers, being the main bottleneck of untargeted metabolomics studies based on LC-ESI-HRMS.

5. Data processing

Huge amounts of data are generated in untargeted metabolomics using chromatographic techniques coupled to HRMS (xC-HRMS). The objective of the data processing is to extract the information of the detected features from the xC-HRMS raw 3D data and obtain a 2D data matrix where they are characterized by *m/z* ratio, retention time (RT) and their relative intensities across the samples, which will be used for statistical analysis. The main steps are (i) *Peak picking and deconvolution*. It consists on the detection of each measured ion in a sample and the assignation to a feature (*m/z* and RT). The peak picking algorithm and deconvolution works with the extracted ion chromatograms attending to some parameter of maximum *m/z* error, interval of time (minimum and maximum time width to be considered a chromatographic peak) and the minimum height or intensity, signal to noise ratio (S/N), among other parameters. (ii) *Retention time alignment*. The matched peaks with similar retention times and *m/z* ratio across multiple samples are grouped in accordance to a window of *m/z* and RT, to be assigned as the same feature and subsequently aligned. This parameter is especially important in LC, as it tends to present more drift that can cause slights differences in retention times along the run. The grouped peaks are then integrated and a peak height or peak area is assigned to the feature in each sample. (iii) *Gap filling*: It is applied to correct and fill in the missing peaks (0 signal) or peaks not detected due to the restrictions of the first two steps

(lower intensities or bad peak shape in some of the samples), but actually may be present, that can affect the power of subsequent statistical analysis.

At this point, a first data table is obtained, and the quality of the features data have to be assured in order to refine the data matrix. Some methodologies as normalisation, scaling and data transformation allow the removal of unwanted variabilities that occur due to both experimental (systematic human and instrumental errors during the analytical process) and biological (e.g. number and size of cells, concentration of biofluids…). These corrections can be grouped as method-driven (normalisation based on internal/external standards and/or quality control samples) and data-driven (scaling and data transformation) [68]. The different approaches that can be applied depending on the source of variability will not be discussed here. As a reference, the reviews from Dudzik et al., describing strategies for quality assurance in the hole untargeted metabolomic process [69], and from Li et al. about different refining methodologies [70] can be consulted. Nevertheless, it is worth noticing the need to include quality control samples (QCs, representative average sample formed by a pool of all samples analysed) in the metabolomic run (e.g. every 5 or 10 samples), to monitor the instrumental analysis, and for validating the features in the data matrix [71]. This surveillance could be applied for example: 1) to filter those features absent in a certain number of QC samples [39,72]; 2) correct intensity drifts caused by variations during the analysis, a common method is to apply the locally estimated smoothing function (LOESS) [26,73]; and 3) to determine the repeatability of each feature along the QCs, removing from the data matrix those with high relative standard deviation (% RSD) [31,74].There is a wide range of informatics tools to perform this important part of the untargeted metabolomic process for xC-HRMS data.

Whether they are free or commercial tools, the processing is mainly the same, although it may differ in how the steps are carried out, some of them working with in-house made algorithms. In the literature reviewed, the most employed tools were open-source software as XCMS (R package or Online) [17,18,75,76], MZmine [16,77] and MetAlign [78]; and commercial software as Mass Profiler Pro (Agilent Technologies Inc.) [79,80], Progenesis QI (Non Linear Dynamics, Waters) [15,25,72], MarkerLynx (Waters) [81,82], SIEVE

(Thermo Scientific) [37] and Compound Discoverer (Thermo Scientific) [83]. Some of them, as the open source MS-DIAL [84] or Compound Discoverer (Thermo Scientific) among others commercial tools, not only perform the abovementioned processes, but also the extraction and deconvolution of DDA and DIA spectral data and annotation by comparison of the deconvoluted MS/MS spectra with in house and/or public data bases, which is especially important for conventional DIA spectral data interpretation [49,85,86]. Because the GC-HRMS technique is less used in this area, most of the listed tools were developed for LC-HRMS data. However, their application to the GC-HRMS data appears to be equally powerful. The tools used were basically proprietary software as Chroma TOF (LECO) [27] or MetaboScape (Bruker Daltonik) [50], and freeware as MzMine [11,19] or MS-DIAL [29]. There are other tools that are gathering strength due to the good results obtained in this type of data, such as PARAFAC2 based Deconvolution and Identification System (PARADISe) [87].

6. Statistical analysis

Although the aim when analysing data from foodomics studies seems quite simple: find the differences in the metabolite profiles related to the experimental design, the complexity and size of the data, the elevated number of metabolites and the natural biological variation of individuals make challenging this exploration. Multivariate data analysis is a powerful tool to explore correlations or co-variations in such datasets. This can be done with (supervised) or without (unsupervised) a priori knowledge about the experimental design [88]. Different tools have evolved during the past few years, but the most often used chemometric method in unsupervised analysis includes principal component analysis (PCA).

Many software options, free and commercial, are available for univariate and/or multivariate statistical tests. Among commercial software, SIMCA P+ (Soft independent modelling by class analogy) (Umetrics, Sweden) [12,23,25,28,47,[89], [90], [91]] is one of the most used, as well as its light version EzInfo, (Umetrics, Sweden) [11,92]. Regarding the free software, MetaboAnalyst, which also provides a companion R package (MetaboAnalystR) to complement the web-based application, is a suitable option [57,86].

There are several aspects to consider before facing the modelling of metabolomics data where the number of variables largely exceeds the number of objects. Data cleaning by one or successive pre-filtration steps should be able to reduce the number of features and eliminate irrelevant signals while avoiding or minimizing relevant chemical information loss. The most commonly used are: i) removal of variables that exhibit a poor stability, meaning relative standard deviation (% RSD or CV %) on peak area across the QCs [91]; ii) removal of variables not present in a minimum number of the samples in one group; iii) removal of those variables that have zeros in a determined number of the samples (if it applies) (retain features with "nonzero values") [93]; iv) removal of variables that show a low fold change or no significant difference among sample groups or among blank runs and any of the sample groups [75,78,94]. Multiple univariate analysis tools as pairwise t-test, ANOVA, etc are available at this point to determine significant differences. With one or more of these pre-filtration steps, a more robust dataset is obtained with still sufficient markers to enable a meaningful analysis. An improvement in the clusterisation of the samples is also observed, with a tight clustering of the QC samples and an increase in the explained variance [10,83].

Once dataset is pre-filtered, PCA can be applied as a first step for interrogating the data in order to observe trends, grouping and/or outliers. PCA obtains new uncorrelated variables preforming linear combinations of the original ones, called principal components (PCs), according to common patterns and maximizing the variance in data. In this way, the dimensionality of the data is reduced while still preserving information from the original data set. The first PC represents the largest variation in the data set. The second PC, orthogonal to the first, covers as much of the residual variation as possible, and so on. Objects far apart in the score plot are different with respect to what patterns the model describes and objects in close proximity exhibit similar variations (see **Fig. 2A**) [95].

In some cases, PCA is enough to determine if the classes can be predicted from the variables (discriminatory PCA) and to identify which ones are important in predicting class membership. PCA allowed the identification of markers potentially useful for the detection fraudulent use of chicken "dead

on arrival" instead of normally slaughtered ones [10]. It was also successfully applied when discriminating between three different studied legumes in order to fight against food fraud [26]. PCA has also been used as exploratory tool previous to supervised analysis for gaining an in-depth understanding of the inherent differences among samples. In this line, PCA was able to suggest that metabolomic changes during milk fermentation by *L. helveticus H9* were more obvious at the fermentation phases (0–8 h), as PCA scores of earlier time points scattered away from those of the later time points (beyond 10 h) and this information was useful for further supervised analysis [92].

Unsupervised hierarchical cluster analysis (HCA), with a heatmap plot, can be used also as exploratory method to observe clusters, analyse and visualize the metabolome differences and/or to confirm the classification performed by PCA [26,56,78]. As an example, HCA could distinguish 5 main groups of metabolites among the 282 serum metabolites after the intake of milk and yogurt; 236 metabolites increased postprandially and 46 features decreased postprandially [72].

Supervised techniques can be very helpful for highlighting sample/group differences when PCA results are masked by high levels of spectral noise, strong batch effects, or high within group variation among other reasons. Partial least squares-discriminant analysis (PLS-DA) is a supervised method that uses multiple linear regression to find the direction of maximum covariance between a data set and class labels. PLS-DA highlights the separation between groups of observations and identifies variables that have most of the class separating information. As an example, the clustering of malt and blended whiskies previously observed in PCA was subsequently highlighted by the PLS-DA indicating that highly significant differences exist among the two Scotch Whisky categories [33]. A variant of PLS-DA is orthogonal partial least squares-discriminant analysis (OPLS-DA), where the variation in the data is divided in between-classes and within-classes (predictive and non-predictive) that are forced to be described, by the first and second OPLS-DA component. Although it does not alter the performance of the classification model of a PLS-DA, it has an easier interpretation [96]. As an example, OPLS-DA was used to develop a model

enable to differentiate between no red meat intake and high red meat intake, in serum samples [15].

However, there is no guarantee that the main variation extracted by the PCA is reflecting the hypothesis put forward. PCA analysis failed to separate samples based on the production system but highlighted the potential effect of the production year on a carrot metabolome study (**Fig. 2A**). The data was then subjected to OPLS-DA and the model was refined (**Fig. 2B**). Variables that contributed to the classification of samples based on production year were investigated and removed from the datasets. This was crucial to improve predictive ability, specificity and sensitivity values of the models [28].

Different methods exist to perform the selection of markers. From the PCA it can be done using loading plot, the backbone of the PCA model. From the loading plot of PCA it was possible to find out which metabolites mainly contribute to the separation of licorice samples from three different origins and species [78]. From PLS-DA, Variable Influence on the Projection (VIP) values > 1−2 generally represent those metabolites carrying the most relevant information for class discrimination. From the OPLS-DA, a combination of VIP and *p(corr)* derived from the S-plot, is a good strategy to identify metabolites with the highest influence on the group separation. VIP >1.0 and *p(corr)* > 0.5 cut-off allowed the selection of most relevant metabolites detected in liver of Wistar rats for the separation of the high-cholesterol (HC) and high-cholesterol enriched with onion (HCO) feeding groups [90]. There are still few studies that only use univariate analysis for discrimination where a wide number of different tests can be found [10]. However, a combination of outputs coming from univariate and multivariate analysis is the most satisfactory and complete strategy for selection of markers [29,97]. Regarding univariate analysis normally used, they can be divided among pair tests (one-way ANOVA, Student's *t*-test, etc) and non-pair tests (Kruskal-Wallis, the Mann-Whitney U test, Welch *t*-test, etc) depending on the normality of the data [29,36,57,79]. These tests should be followed by a False Discovery Rate calculation p-FDR < 0.05 (q value set at 0.01) normally applying Benjamini-Hochberg procedure to rectify p-values in order to correct for multiple hypothesis testing and reduce the false positives than are

susceptible to occurred when the number of variables largely exceeds de number of objects [36,47,91].

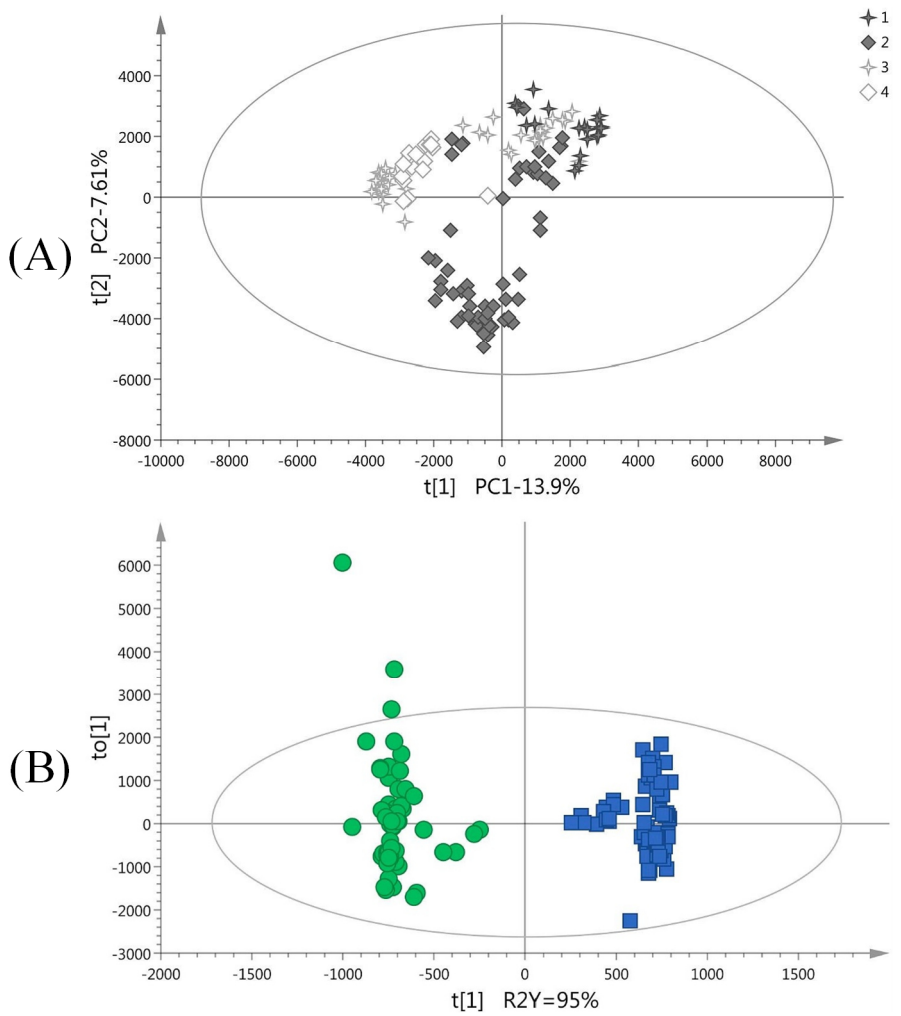


Fig. 2. *Figure constructed from Cubero-Leon et al. [28]. (A) PCA score plot where the first and the second principal components (t1 and t2) are shown. Each harvested year is represented with a different symbol. In picture legend 1: year 2005; 2: year 2006; 3: year 2007; 4: year 2008. (B) Score plot of OPLS-DA of model 10. The first predictive component (t1) and the first orthogonal component (to1) are shown. R2Y: explained variation. Ellipse Hotelling's T2 (95 %). Organic samples (filled circles), conventional samples (filled squares).*

As an example, combination of VIP>1, *p(corr)*≥0.5, with fold-change≥1.5 and p-value <0.05 (one-way ANOVA) was used for biomarkers selection of discriminant macrophages metabolites between control and high-dose group of *Panax ginseng* group [97]. Two-way ANOVA analysis is a potent approach that can be added to discover the metabolites affected by two factors. As an example, the level of *Lonicerae Japonicae Flos* and the administration days (time) were the two factors that affected the metabolism of the rat [74].

The quality of the models is generally evaluated by the goodness-of-fit parameter (R2X), the proportion of the variance of the response variable that is explained by the model (R2Y) and the predictive ability parameter Q2. R2X, R2Y and Q2 values close to 1 indicate an excellent model, and thus values higher than 0.5 indicate good quality of PLS-DA and OPLS-DA models. However, it is remarkable that, contrary to PCA, these supervised methods tend to overfit models and can generate excellent class separation even with random data. For this reason, results of these types of tests should be critically checked and properly cross-validated using procedures in which some of the samples are left out and their classification have to be predicted. In order to test for possible overfitting and to confirm that Q2 values are stable and relevant, permutation tests are used. As an example, 7-fold full cross-validation and permutation test on the responses (500 random permutations) were performed, in order to avoid over-fitting and prove the robustness of the obtained models [81]. Some authors works claim that findings need to be further verified using a higher number of samples, other statistical tools and/or other analytical tools [56]. Apart from the classification rates obtained by internal cross-validation (automatically performed by software like SIMCA P+), external validations using samples that had not been used for the construction of the models is not such a frequent practise but definitely adds value to the developed model [25,28]. Sales et al. even performed a reduction of variables until 15 to create and validate a model that could be used as starting point for classification of future olive oil samples by quality following a simpler targeted analysis [11]. Chatterjee et al. developed an LC-MS/MS assay with the set of 34 markers identified for rapid authentication of shrimps species and it was tested with

unknown shrimp samples from the market [25]. In the field of biomarkers of food intake (BFI), an independent separate controlled, single-blinded, cross-over meal study was carried out to validate the candidate biomarkers of meat intake identified in a previous study resulting in a set of six better validated candidate markers that were further used to predict beef intake [16].

Another important aspect of the validation of biomarkers is the biological plausibility of such identified makers. Additionally, in the field of BFIs, examination of dose-response has become an essential prerequisite to demonstrate the use of biomarkers in dietary assessment for further applications in nutritional epidemiology. Subsequent confirmation and validation of biomarkers in intervention, independent studies, other cohorts, less-controlled, also adds evidence to the output [16,36,54].

7. (Bio)markers identification

Structural characterization and elucidation of potential markers highlighted in the statistical data analysis is commonly a challenge in metabolomics and can become the bottleneck of the overall metabolomics process. In HRMS, accurate mass measurement is the gold-standard for identification procedure and it is essential for facing this process. Q-TOF and OT-based HRMS analysers are more and more popular because of their high specificity, high resolution and low exact mass deviation [57]. The current methods and tools available for annotation of metabolites in untargeted metabolomics studies applying LC-MS platforms have been recently reviewed [98].

Chemical Analysis Working Group, within Metabolomics Standards Initiative (MSI), proposed four levels of confidence in metabolite identification: Level I is for identified/confirmed compounds, when their identity is validated using authentic standards and subsequent MS analysis; Level II is for putatively annotated compounds (e.g. without chemical reference standards, based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries)); Level III is for putatively characterised compound classes (e.g. based upon characteristic physicochemical properties of a chemical class of compounds, or by spectral similarity to known compounds of a chemical class); Level IV is for unknown

compounds —although unidentified or unclassified, these metabolites can still be differentiated and quantified based upon spectral data [99]. More recently Schymanski et al. have reported a similar system but including five-levels of confidence [100]. It has been updated including ion mobility separation as an additional parameter for more reliable identifications [101].

The first step in the identification workflow is to recognise the (quasi)-molecular ion in the accurate mass spectrum (typically, protonated or deprotonated molecule in LC-MS), where the presence of adducts must be also taken into account. Some tools, like CAMERA for XCMS or Progenesis QI, allow componentization, which means that different signals from the same metabolite are grouped together offering greater confidence to the annotation. Then, the most likely elemental composition is calculated according to the mass error and isotope pattern. After that, fragment ion information based on MS/MS or DIA experiments is used to establish the fragmentation pathways and discard possible chemical structures. To this aim, the use of offline/online and commercially/freely/in-house available spectral databases, are of great help [102]. The most used databases in the reviewed literature are: METLIN repository database (*https://metlin.scripps.edu*), Human Metabolome Database (HMDB) (*http://www.hmdb.ca*), Kyoto Encyclopedia of Genes and Genomes (*https://www.genome.jp/kegg/*), FooDB (*http://foodb.ca*), Chemspider (*http://www.chemspider.com/*), PubChem Compound database, LIPID Metabolites and Pathways Strategy LIPID MAPS (*https://www.lipidmaps.org/*), Chemical Entities of Biological Interest (ChEBI), SIRIUS, CSI:Finger ID. In-silico fragmentation tools are also useful in this process, emphasizing MetFrag *(https://msbi.ipb-halle.de/MetFrag/*), MetFusion, MassBank (*http://www.massbank.jp*), FooDB (*http://foodb.ca*) and Competitive Fragmentation Modelling for metabolite identification (CFM-ID) (*http://cfmid.wishartlab.com/*), among others.

Several software programs are available to automatize and simplify this challenging process: Progenesis QI, Compound Discoverer, MS-DIAL, CSI: FingerID, and MyCompoundID. However, the expertise and knowledge of the analyst on mass spectrometry and fragmentation rules is crucial to avoid false identifications. The injection of a reference standard, if commercially

available, is the last step to assure the identity of the marker. When not available, the synthesis of the candidate compound may be required for full confirmation of the identity. Once the markers are identified, quantitative methods using standard substances can be developed to confirm that the specific markers do accurately reflect the differences between the classes. As an example of this identification workflow, chlorogenic acid was highlighted as potential biomarker for Colombian coffees discrimination [48] according with the annotation performed after comparison with Metlin. Additional MS/MS experiments allowed the fragmentation evaluation for structure confirmation with the aid of in silico tools like Mefrag (**Fig. 3**).

The availability of NIST library is a clear advantage for marker identification in metabolomics studies based on GC-EI-MS analysis. Although this library works mainly in nominal mass, a first step in the identification of metabolites is possible. It can be supported by the isotopic pattern, exact mass for parent ion (if exists) and fragments, and Kovats retention index. Soft ionization sources as positive chemical ionization (PCI) or atmospheric pressure chemical ionization (APCI) enable obtaining highly diagnostic molecular ions and/or protonated molecules of compounds which are extensively fragmented under EI conditions [11,33]. Compound identification by GC–MS can also be complemented using FiehnLib library and the Golm Metabolome Database.

Unfortunately, despite the efforts invested in biomarker identification, this final goal is not always achieved. However, the minimal requirements of reporting for unknown metabolites (retention time, prominent ion and fragment ion) can still be fulfilled [99]. This was the case of the study by Chatterjee et al., in which some markers only yielded one fragment ion, thus decreasing the reliability of the identification [25].

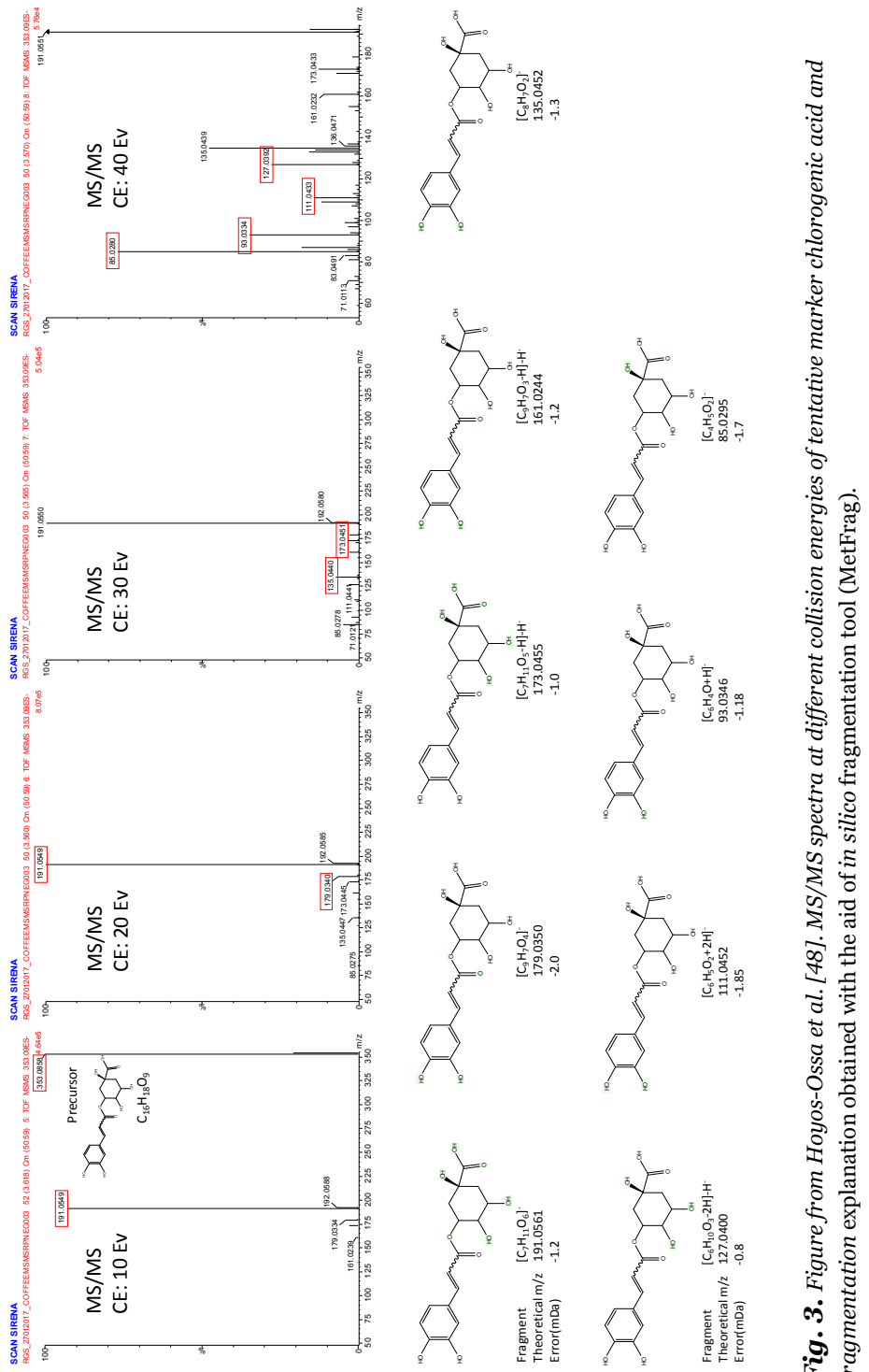


Fig. 3. *Figure from Hoyos-Ossa et al. [48]. MS/MS spectra at different collision energies of tentative marker chlorogenic acid and fragmentation* explanation obtained with the aid of *in silico* fragmentation tool (MetFrag).

8. Applications

In this section, we outline a selection of untargeted metabolomic studies that made use of LC and/or GC coupled to HRMS in the field of food processing, including authenticity, quality and safety [4,103]; the discovery of biomarkers of food intake [[104], [105], [106]]; and assessment of effects of food and diet on health [[107], [108], [109]].

8.1. Food processing

The growing demands by our society, authorities and scientists to advance knowledge about the food consumed has led to the development of robust analytical methodologies to improve the quality and safety of food products and prevent food frauds. Several applications have been developed in the last years related to food authenticity. Some works were directed towards the identification of markers for characterization of food samples by its geographical origin in honey [57], garlic [14] or Adzuki Bean [30]; and for authentication of Protected Designation of Origin (PDO) of Grana Padano cheeses [56] and Colombian coffees [48]. Moreover, patterns of different agricultural practises were assessed in carrots [28], different varieties of legumes [26], potato [110], *Vaccinium* fruits [94] and almonds [111], as well as different species identity, geographical origin and production method of commercially prawn and shrimps [25]. The characterization of food products by untargeted approach was also determinant to prevent fraudulent practises, such as the production of adulterated fruit juices in citrus [89], the production of counterfeited Scotch Whisky [33], and dead on arrival instead of the typical slathered poultry meat [10]. The characterization of organic culture practises against traditional cropping systems in wheat grains [23] has been also evaluated.

The characteristics of food appreciated by customers, including appearance, texture, flavour, aroma and nutritional composition, are also crucial in food quality, and are often dependent on subtle changes in the food's metabolome [4]. Regarding the appearance of food, it was established a relation between the colour of sea buckthorn and its chemical properties, having the red ones a better quality [13]. In other experience, the aim was to discover biomarkers related to the taste of food as it was the case with quality

assessment of green tea [12] and olive oil [11]. Another important factor of food quality is to find markers related to storage time since the food quality worsens. Thus, significant differences in the metabolites composition of chilled chicken meat were found in accordance with conservation period [24].

Regarding food safety, there is a demand of robust markers for prevention of bad practises and possible errors in the food supply chain [4]. In this line, new markers of egg ageing were found by an untargeted metabolic approach [83]. Regarding bad agricultural practises, a strategy was developed to discriminate green tea samples in concerning their contamination levels [75]. To ensure food safety and quality, it is important food traceability, which means continuous monitoring of the foods products through the entire supply chain, enabling the correction of mistakes. The role of HRMS-untargeted metabolomics in this context is the identification of characteristics markers of each stage of the process. In this way, metabolomics was found a powerful tool to identify different patterns between fresh tiger nut milk and milk processed by ultra-high temperature treatment [91]. It was also employed for the investigation of potential markers of three different species of licore plants (*Glycyrrhiza species*), which are sweetening and flavouring agents in food and beer industries [78].

8.2. Food intake

A significant challenge in nutritional research is the measurement of dietary intake, which must be both accurate and applied to large numbers of people [112]. Traditionally, FFQ, 24-h recall or other dietary assessment tools, have been the standard tools for dietary assessment. Unfortunately, these approaches are subjected to errors such as underreporting, recall errors and difficulty in assessment of portion sizes, which generate biased and inaccurate results and associations. The scenery in nutritional epidemiology changed with the emergence of high-throughput metabolomics techniques enabling the discovery of novel biomarkers of food intake (BFIs) that represent objective measures of dietary and specific food intake.

In the literature, several examples can be found on intervention studies that employ untargeted chromatography coupled to HRMS-based metabolomics for the discovery of food-derived metabolites in banana [55],

pea [54], fermented dairy products [72], different varieties of tomatoes [79] and tomato juice [77]. Biomarkers related to coffee consumption habits in various European countries were researched in a cross-sectional study within the European Prospective Investigation on Cancer and Nutrition (EPIC) [113]. In another observational study within the Singapore Prospective Study Program (SP2), patterns of meat and seafood consumption were assessed based on plasma metabolic profiles [19]. Regarding meat consumption, a great interest exists in finding indicators of red and processed meat intake since its consumption is associated with the development of chronic diseases [15,16,20,36].

The identification of biomarkers related to the intake of supplements suspected of having a benefit for human health is also another field of recent research. Several interventional studies with different bioactive foods and supplements have been performed: bioactive garlic [38], kiwi wine [27], beetroot juice [61], *angelica keiskei* [60], green coffee bean extract (GCBE) [81] and *amalaki rasayana* [39]. The metabolic patterns related with food enriched with some bioactive compound have been also investigated, as for example, flavan-3-ol-enriched dark chocolate, compared with standard dark chocolate and white chocolate [17] and apple juice enriched with four groups of polyphenols [37].

A better understanding of the relation between dietary patterns and metabolic profiles is crucial for improving the recommendations of health authorities about what diet is better for a better quality of life. New Nordic diet (NDD), which was designed to be balanced and healthy, was compared to average Danish diet (ADD) in a long intervention study, identifying potential metabolic patterns that indicate potential health benefits of the NDD [114]. On the other hand, a detailed dietary assessed method was employed in the *Supplémentation en Vitamines et Minéraux AntioXidants* (SU.VI.MAX) cohort with the aim of performing a cross-sectional study and look at the difference in the plasma metabolic profiles according to their adherence to the French dietary recommendations [18].

As supported by several studies, HRMS-based untargeted metabolomics is a powerful approach in the discovery of new BFIs. Per definition, metabolomics fingerprinting is a data-driven approach, what

means that a new hypothesis is forged from the biomarkers discovered. Therefore, all BFIs discovered by untargeted approach are tagged as "putative" since its necessary a proper validation process to confirm the association of robust BFIs to a specific food or diet. BFIs discovered in intervention studies used to be confirmed by the use of independent cohort studies (cross-sectional), as for example, Karlsruhe Metabolomics and Nutrition (KarMeN) [55] and EPIC study cohorts [20,36]. Other strategies include the use of dose-response for validation and independent study for confirmation [54]. In the case of potential BFIs identified only in cohort studies, these do not assess a correlation with the food consumed but rather an association and should be confirmed with an interventional study to validate them [112]. Nevertheless, as there is not an established standard methodology for validation of BFIs, L.O. Dragsted et al. proposed validation criteria based on analytical and biological aspects [115].

8.3. Food and health effects

Since metabolomics can provide a complete picture of the general dietary intake and reflect the current biological status of an individual, another goal of untargeted metabolomics in the nutrition field is to study the complex relationships between nutritional exposure and the positives or negatives effects on health/disease state [105,116]. The information obtained not only allows an accurate monitoring of a diet and lifestyle but may also help to design strategies to manipulate the physiological state with the ultimate goal to improve the individual health thought personalised dietary interventions [108,109].

Untargeted metabolomics approaches based on chromatography-HRMS have been applied to determine how a whole diet can affect the health state and to identify the molecular mechanism involved [41]. To this aim, both interventional, with human or mice/rat models, and observational studies have been carried out. The objective was to determine the changes occurred in the metabolism under a specific diet [66] or the differences obtained between 5 diets in different mice tissues [117]. Showalter et al. performed a multiplatform untargeted metabolomics study [49], finding significant metabolic alterations that suggest that the physiology of lungs can be altered by obesity. In other studies, the goal was to determine the

relationship between the diet and a specific disease. For example, the use of multifunctional diet in order to reduce cardiometabolic risk factors [47], the adherence to a healthy Nordic diet of a Swedish prospective cohort and the risk of future type 2 diabetes [65]; the potential of ketogenic diet as an auxiliary cancer therapy with tumour Xenograft mouse models [46]; and the correlation of diet with microbiota and metabolism of inflammatory bowel disease human patients [118]. Given the diversity and complexity of diet constituents, some studies were focused on one diet constituent and the effects on health, as fish or coffee intake and type 2 diabetes risk [119,120]; or the health detriment due to the consumption of heated soybean oil [121] or sweetened beverages [73].

There is a growing interest in nutraceuticals or diet supplements, especially in the so called 'functional foods', food products to which a health benefit is attributed (naturally or artificially added) besides its own nutritional contribution [122,123]. Nevertheless, a wide variety of food products are potentially beneficial for health and it might be difficult to determine if they can be classified as 'functional'. For this reason, a notable number of untargeted metabolomic approaches have been performed to determine the impact on health or disease of specific products considered as functional foods: ginseng [82,97], herbal traditional medicines [59,74],wholegrain rye bread [80], walnuts [85], lettuce [58] and onion [90]. The controlled trial study of fish oil supplementation during pregnancy, ended in the detection of several altered metabolic pathways significantly associated with a reduced risk of asthma by age 5 [124]. Likewise, the effects of diets supplements were assessed, as selenium impact on metabolic disorders [93] or the use of xenoestrogens in combination with cancer therapy [76].

Moreover, it is important the characterization of food products for their validation as functional food and in order to enhance their potential. An untargeted lipidomic approach was applied for the discovery of potentially high valuable polar lipids of *Porphyra dioica*, algae commonly used for *sushi* preparation [63]. HRMS-based metabolomics was also used to study the process of probiotic food product process as the dynamics of skim milk

fermentation by *L. helveticus H9* strain [92] and the distinction between the biofilm and planktonic state *B. bifidum* strain [29].

It is worth noticing that the term functional food is usually applied to food products that have naturally or artificially substances known for their benefits on health, such as essential fatty acids, flavonoids, vitamins, polyphenols, etc. [122]. For this reason, determining whether a food is a functional product is often carried out through target analysis of the compounds that are known to have a beneficial effect. There is abundant bibliography available on targeted metabolomics in this field, which however does not fall within the scope of this review. It should be noted that target and untargeted metabolomics approaches can be combined, as for example to study the effects of white-blue light and dark in growth of cacao cell suspensions [86].

9. Future prospects and conclusions

The use of MS-based approaches for untargeted metabolomics for investigation of food (bio)markers is still far from reaching its maximum potential. HRMS will surely be dominant in the near future, and the continuous improvements in instrumentation will be translated to enhanced capabilities of the developed strategies. For example, to maximize the metabolome coverage, it is necessary to acquire MS data in complementary chromatographic and ionization modes, but also MS/MS data, which can be acquired under DDA and/or DIA modes with sequential mass windows (e.g., SWATH, SONAR).

Regarding DDA, however, the metabolite coverage is not usually enough, and many features may lack MS/MS data. Technological evolution has allowed improvements in this acquisition mode and increasing the acquisition speed, which together with new software developments make possible to perform automated and iterative DDA in the newest instruments [125]. This strategy automates iterative exclusion and inclusion lists to reduce the fragmentation of redundant features coming from the background and allows performing exhaustive precursor selection obtaining more relevant MS^2 spectra. Such lists are automatically imported into the DDA method before the first ddMS^2 acquisition of the sample and are updated

prior the next ddMS^2 run, bypassing precursors already fragmented to the exclusion list. MS^2 spectra are acquired for compounds remaining on the inclusion list. This approach enables to cover a wider range of compounds (including the lower-abundance ones) that were lost by the traditional DDA methods. However, in order to not increase significantly the acquisition time, such strategy is only applied to the QC samples, as it requires multiple reinjection until reaching the complete coverage of the compounds. In this way, the samples are acquired in FS mode, and the iterative DDA is only applied to a reduced number of QC samples for future compound characterization.

In relation to DIA, the incorporation of ion mobility spectrometry (IMS) to HRMS has allowed a new DIA mode. An example is High Definition MS^E (HDMS^E). As occurs in conventional MS^E, two functions are acquired at low and high collision energy, but after ion mobility separation. In this way, the precursor and the product ions are recorded with the same drift time (translated into CCS, Å^2). This opens the possibility to filter the fragmentation spectra (obtained from all the ions fragmented in the scan cycle) by the drift time of a target ion and to obtain cleaner spectra without interfering fragments of co-eluting ions. Thus, the visualization of only the products ions from a specific precursor is feasible, enhancing the purity of the MS^2 spectrum with the inherent benefits of DIA acquisitions regarding available MS/MS data for all future biomarkers. The potential of this technique has been recently evaluated for orange dietary biomarkers discovery [64] and implemented for comparison of different polar lipids extraction methods to be used in evaluation of botanical origin, with potatoes as a case of study [110]. In both studies, data processing was performed using Progenesis QI (Waters), a unique software, able of performing the processing of 4D data obtained with xC-IMS-HRMS instruments.

In terms of ionization techniques, the recent atmospheric pressure CI source (APCI) is an attractive alternative to EI in GC-HRMS analysis. APCI enables a soft ionization ensuring the preservation of the (pseudo)-molecular ion, which is of great interest when the molecular ion is absent from the highly fragmented EI spectrum, which would imply a reduction in the selectivity and sensitivity. As APCI works under atmospheric pressure, the same mass

analyser can be shared by both LC and GC instruments, since the vacuum does not need to be broken as occurs with EI and CI sources [126]. However, due to the novelty of this technique, there is a lack of spectral databases under this ionization source in comparison with EI. Only two articles using GC-APCI-HRMS have been found, both related to olive oil. Sales et al. studied the volatile composition of olive oil to develop a classification model for quality assessment [11], and Olmo-García et al. applied this technique for olive oil origin discrimination [50].

All in all, the combination of gas and liquid chromatography with high-resolution mass spectrometry, together with technological advances in instrumentation, both in chromatography (e.g. new stationary phases, format and particle size) and HRMS (e.g. resolution power, acquisition speed, MS^2 acquisition modes) have been crucial to explain the impulse of untargeted metabolomics in the last few years. In particular, this approach has driven the expansion of knowledge on food processing, intake and the effects of food in health. The hyphenation of modern chromatography and HRMS allows a highly efficient separation combined with the acquisition of sensitive and high-quality structural compound information, facilitating the detection and identification of metabolites in complex biological samples, such as food matrices or biofluids. For this reason, this hyphenation has become one of the most used techniques in untargeted metabolomics studies in the field of food and nutrition. The implementation of chromatography-HRMS techniques, together with correct study designs and appropriate sample treatments, as well as the use of upgraded data treatment programs and powerful statistical tools, has notably enhanced the capabilities of untargeted metabolomics in the food field.

The increasing demand for more exhaustive control over food processing, in terms of authenticity, quality and safety, can be met, addressing needs such as the characterization of food products by geographical origin or production method, and the detection of adulteration or bad practises. Regarding nutrition, the application of untargeted metabolomics using chromatography-HRMS has revealed potential biomarkers related to the intake of food products and diets. Moreover, this approach can help to understand the complex relationships between

nutritional exposure and physiological state, by the study of the effects of diet, or potentially beneficial food products, on the metabolism, as well as to evaluate the benefits to health.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

L Lacalle-Bergeron acknowledges the financial support of Universitat Jaume I, Spain for his pre-doctoral grant (UJI 19I001/03). T. Portolés acknowledges Ramon y Cajal Program from the Ministry of Economy and Competitiveness of Spain (RYC-2017-22525) for funding her research. D. Izquierdo-Sandoval acknowledges the Ministry of Science, Innovation and Universities of Spain for funding his research through the FPU pre-doctoral program (FPU19/01839). The authors acknowledge the financial support of Universitat Jaume I (UJI-B2020-37) and Generalitat Valenciana, as research group of excellence PROMETEO/2019/040.

References

[1] M. Herrero, C. Simó, V. García-Cañas, E. Ibáñez, A. Cifuentes, Foodomics: MS-based strategies in modern food science and nutrition, Mass Spectrom. Rev., 31 (2012), pp. 49-69, 10.1002/mas.20335

[2] J.-L. Wolfender, G. Marti, A. Thomas, S. Bertrand, Current approaches and challenges for the metabolite profiling of complex natural extracts, J. Chromatogr. A, 1382 (2015), pp. 136-164, 10.1016/j.chroma.2014.10.091

[3] P. Maruvada, J.W. Lampe, D.S. Wishart, D. Barupal, D.N. Chester, D. Dodd, Y. Djoumbou-Feunang, P.C. Dorrestein, L.O. Dragsted, J. Draper, L.C. Duffy, J.T. Dwyer, N.J. Emenaker, O. Fiehn, R.E. Gerszten, F.B. Hu, R.W. Karp, D.M. Klurfeld, M.R. Laughlin, A.R. Little, C.J. Lynch, S.C. Moore, H.L. Nicastro, D.M. O'Brien, J.M. Ordovás, S.K. Osganian, M. Playdon, R. Prentice, D. Raftery, N. Reisdorph, H.M. Roche, S.A. Ross, S. Sang, A. Scalbert, P.R. Srinivas, S.H. Zeisel, Perspective: dietary biomarkers

of intake and exposure - exploration with omics approaches, Adv. Nutr., 11 (2020), pp. 200-215, 10.1093/advances/nmz075

[4] M. Castro-Puyana, R. Pérez-Míguez, L. Montero, M. Herrero, Application of mass spectrometry-based metabolomics approaches for food safety, quality and traceability, TrAC Trends Anal. Chem., 93 (2017), pp. 102-118, 10.1016/j.trac.2017.05.004

[5] K. Dettmer, P.A. Aronov, B.D. Hammock, Mass spectrometry-based metabolomics, Mass Spectrom. Rev., 26 (2007), pp. 51-78, 10.1002/mas.20108

[6] A.E. Lytou, E.Z. Panagou, G.-J.E. Nychas, Volatilomics for food quality and authentication, Curr. Opin. Food Sci., 28 (2019), pp. 88-95, 10.1016/j.cofs.2019.10.003

[7] F. Vivanco, M.G. Barderas, C.M. Laborde, M. Posada, F. De La Cuesta, I. Zubiri, G. Alvarez-Llamas, Metabolomic profiling for identification of novel potential biomarkers in cardiovascular diseases, J. Biomed. Biotechnol., 2011 (2011), pp. 1-9, 10.1155/2011/790132

[8] H.G. Gika, G.A. Theodoridis, R.S. Plumb, I.D. Wilson, Current practice of liquid chromatography–mass spectrometry in metabolomics and metabonomics, J. Pharmaceut. Biomed. Anal., 87 (2014), pp. 12-25, 10.1016/j.jpba.2013.06.032

[9] W.B. Dunn, I.D. Wilson, A.W. Nicholls, D. Broadhurst, The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans, Bioanalysis, 4 (2012), pp. 2249-2264, 10.4155/bio.12.204

[10] K.L. Sidwick, A.E. Johnson, C.D. Adam, L. Pereira, D.F. Thompson, Use of liquid chromatography quadrupole time-of-flight mass spectrometry and metabonomic profiling to differentiate between normally slaughtered and dead on arrival poultry meat, Anal. Chem., 89 (2017), pp. 12131-12136, 10.1021/acs.analchem.7b02749

[11] C. Sales, M.I. Cervera, R. Gil, T. Portolés, E. Pitarch, J. Beltran, Quality classification of Spanish olive oils by untargeted gas chromatography coupled to hybrid quadrupole-time of flight mass

spectrometry with atmospheric pressure chemical ionization and metabolomics-based statistical approach, Food Chem., 216 (2017), pp. 365-373, 10.1016/j.foodchem.2016.08.033

[12] J. Jing, Y. Shi, Q. Zhang, J. Wang, J. Ruan, Prediction of Chinese green tea ranking by metabolite profiling using ultra-performance liquid chromatography–quadrupole time-of-flight mass spectrometry (UPLC–Q-TOF/MS), Food Chem., 221 (2017), pp. 311-316, 10.1016/j.foodchem.2016.10.068

[13] C. He, G. Zhang, J. Zhang, Y. Zeng, J. Liu, Integrated analysis of multiomic data reveals the role of the antioxidant network in the quality of sea buckthorn berry, FASEB J., 31 (2017), pp. 1929-1938, 10.1096/fj.201600974R

[14] V. Hrbek, M. Rektorisova, H. Chmelarova, J. Ovesna, J. Hajslova, Authenticity assessment of garlic using a metabolomic approach based on high resolution mass spectrometry, J. Food Compos. Anal., 67 (2018), pp. 19-28, 10.1016/j.jfca.2017.12.020

[15] D. Carrizo, O.P. Chevallier, J.V. Woodside, S.F. Brennan, M.M. Cantwell, G. Cuskelly, C.T. Elliott, Untargeted metabolomic analysis of human serum samples associated with different levels of red meat consumption: a possible indicator of type 2 diabetes?, Food Chem., 221 (2017), pp. 214-221, 10.1016/j.foodchem.2016.10.056

[16] C. Cuparencu, Å. Rinnan, L.O. Dragsted, Combined markers to assess meat intake—human metabolomic studies of discovery and validation, Mol. Nutr. Food Res., 63 (2019), p. 1900106, 10.1002/mnfr.201900106

[17] L.M. Ostertag, M. Philo, I.J. Colquhoun, H.S. Tapp, S. Saha, G.G. Duthie, E.K. Kemsley, B. de Roos, P.A. Kroon, G. Le Gall, Acute consumption of flavan-3-ol-enriched dark chocolate affects human endogenous metabolism, J. Proteome Res., 16 (2017), pp. 2516-2526, 10.1021/acs.jproteome.7b00089

[18] L. Lécuyer, C. Dalle, P. Micheau, M. Pétéra, D. Centeno, B. Lyan, M. Lagree, P. Galan, S. Hercberg, A. Rossary, A. Demidem, M.-P.

Vasson, V. Partula, M. Deschasaux, B. Srour, P. Latino-Martel, N. Druesne-Pecollo, E. Kesse-Guyot, S. Durand, E. Pujos-Guillot, C. Manach, M. Touvier, Untargeted plasma metabolomic profiles associated with overall diet in women from the SU.VI.MAX cohort, Eur. J. Nutr. (2020), 10.1007/s00394-020-02177-5

[19] Y. Lu, L. Zou, J. Su, E. Tai, C. Whitton, R. van Dam, C. Ong, Meat and seafood consumption in relation to plasma metabolic profiles in a Chinese population: a combined untargeted and targeted metabolomics study, Nutrients, 9 (2017), p. 683, 10.3390/nu9070683

[20] R. Wedekind, P. Keski-Rahkonen, N. Robinot, V. Viallon, P. Ferrari, E. Engel, M.-C. Boutron-Ruault, Y. Mahamat-Saleh, F.R. Mancini, T. Kühn, T. Johnson, H. Boeing, M. Bergmann, A. Karakatsani, A. Trichopoulou, H. Peppa, C. Agnoli, M. Santucci de Magistris, D. Palli, C. Sacerdote, R. Tumino, M.J. Gunter, I. Huybrechts, A. Scalbert, Syringol metabolites as new biomarkers for smoked meat intake, Am. J. Clin. Nutr., 110 (2019), pp. 1424-1433, 10.1093/ajcn/nqz222

[21] J. Pezzatti, J. Boccard, S. Codesido, Y. Gagnebin, A. Joshi, D. Picard, V. González-Ruiz, S. Rudaz, Implementation of liquid chromatography–high resolution mass spectrometry methods for untargeted metabolomic analyses of biological samples: a tutorial, Anal. Chim. Acta, 1105 (2020), pp. 28-44, 10.1016/j.aca.2019.12.062

[22] D. Vuckovic, Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry, Anal. Bioanal. Chem., 403 (2012), pp. 1523-1548, 10.1007/s00216-012-6039-y

[23] Y. Zhang, S. Cao, Z. Zhang, X. Meng, C. Hsiaoping, C. Yin, H. Jiang, S. Wang, Nutritional quality and health risks of wheat grains from organic and conventional cropping systems, Food Chem., 308 (2020), p. 125584, 10.1016/j.foodchem.2019.125584

[24] D. Wen, Y. Liu, Q. Yu, Metabolomic approach to measuring quality of chilled chicken meat during storage, Poultry Sci., 99 (2020), pp. 2543-2554, 10.1016/j.psj.2019.11.070

[25] N.S. Chatterjee, O.P. Chevallier, E. Wielogorska, C. Black, C.T. Elliott

Simultaneous authentication of species identity and geographical origin of shrimps: untargeted metabolomics to recurrent biomarker ions, J. Chromatogr. A, 1599 (2019), pp. 75-84, 10.1016/j.chroma.2019.04.001

[26] R. Llorach, C. Favari, D. Alonso, M. Garcia-Aloy, C. Andres-Lacueva, M. Urpi-Sarda, Comparative metabolite fingerprinting of legumes using LC-MS-based untargeted metabolomics, Food Res. Int., 126 (2019), p. 108666, 10.1016/j.foodres.2019.108666

[27] Q. Zeng, H. Song, X. Xu, W. Mao, H. Xie, J. Liang, X. Chen, D. Chen, Y. Zhan, Health effects of kiwi wine on rats: an untargeted metabolic fingerprint study based on GC-MS/TOF, RSC Adv., 9 (2019), pp. 13797-13807, 10.1039/C9RA02138H

[28] E. Cubero-Leon, O. De Rudder, A. Maquet, Metabolomics for organic food authentication: results from a long-term field study in carrots, Food Chem., 239 (2018), pp. 760-770, 10.1016/j.foodchem.2017.06.161

[29] F.A. Sadiq, B. Yan, J. Zhao, H. Zhang, W. Chen, Untargeted metabolomics reveals metabolic state of Bifidobacterium bifidum in the biofilm and planktonic states, LWT, 118 (2020), p. 108772, 10.1016/j.lwt.2019.108772

[30] T.J. Kim, J.G. Park, S.K. Ahn, K.W. Kim, J. Choi, H.Y. Kim, S.-H. Ha, W.D. Seo, J.K. Kim, Discrimination of *Adzuki* bean (*Vigna angularis*) geographical origin by targeted and non-targeted metabolite profiling with gas chromatography time-of-flight mass spectrometry, Metabolites, 10 (2020), p. 112, 10.3390/metabo10030112

[31] M. Goettel, R. Niessner, D. Mueller, M. Scherer, G. Scherer, N. Pluym, Metabolomic fingerprinting in various body fluids of a diet-controlled clinical smoking cessation study using a validated GCTOF- MS metabolomics platform, J. Proteome Res., 16 (2017), pp. 3491-3503, 10.1021/acs.jproteome.7b00128

[32] A. Erban, I. Fehrle, F. Martinez-Seidel, F. Brigante, A.L. Más, V. Baroni, D. Wunderlin, J. Kopka, Discovery of food identity markers by

metabolomics and machine learning technology, Sci. Rep., 9 (2019), p. 9697, 10.1038/s41598-019-46113-y

[33] M. Stupak, I. Goodall, M. Tomaniova, J. Pulkrabova, J. Hajslova, A novel approach to assess the quality and authenticity of Scotch Whisky based on gas chromatography coupled to high resolution mass spectrometry, Anal. Chim. Acta, 1042 (2018), pp. 60-70, 10.1016/j.aca.2018.09.017

[34] É.A. Souza-Silva, R. Jiang, A. Rodríguez-Lafuente, E. Gionfriddo, J. Pawliszyn, A critical review of the state of the art of solid-phase microextraction of complex matrices I. Environmental analysis, TrAC Trends Anal. Chem., 71 (2015), pp. 224-235, 10.1016/j.trac.2015.04.016

[35] E.J. Want, I.D. Wilson, H. Gika, G. Theodoridis, R.S. Plumb, J. Shockcor, E. Holmes, J.K. Nicholson, Global metabolic profiling procedures for urine using UPLC–MS, Nat. Protoc., 5 (2010), pp. 1005-1018, 10.1038/nprot.2010.50

[36] W. Cheung, P. Keski-Rahkonen, N. Assi, P. Ferrari, H. Freisling, S. Rinaldi, N. Slimani, R. Zamora-Ros, M. Rundle, G. Frost, H. Gibbons, E. Carr, L. Brennan, A.J. Cross, V. Pala, S. Panico, C. Sacerdote, D. Palli, R. Tumino, T. Kühn, R. Kaaks, H. Boeing, A. Floegel, F. Mancini, M.-C. Boutron-Ruault, L. Baglietto, A. Trichopoulou, A. Naska, P. Orfanos, A. Scalbert, A metabolomic study of biomarkers of meat and fish intake, Am. J. Clin. Nutr., 105 (2017), pp. 600-608, 10.3945/ajcn.116.146639

[37] K. Trošt, M.M. Ulaszewska, J. Stanstrup, D. Albanese, C. De Filippo, K.M. Tuohy, F. Natella, C. Scaccini, F. Mattivi, Host: microbiome co-metabolic processing of dietary polyphenols – an acute, single blinded, cross-over study with different doses of apple polyphenols in healthy subjects, Food Res. Int., 112 (2018), pp. 108-128, 10.1016/j.foodres.2018.06.016

[38] Á. Fernández-Ochoa, I. Borrás-Linares, A. Baños, J.D. García-López, E. Guillamón, C. Nuñez-Lechado, R. Quirantes-Piné, A. Segura-Carretero,A fingerprinting metabolomic approach reveals deregulation of endogenous metabolites after the intake of a bioactive garlic supplement, J. Funct. Foods, 49 (2018), pp. 137-145, 10.1016/j.jff.2018.08.003

[39] V. Kumar, A.A. Kumar, V. Joseph, V.M. Dan, A. Jaleel, T.R.S. Kumar, C.C. Kartha, Untargeted metabolomics reveals alterations in metabolites of lipid metabolism and immune pathways in the serum of rats after long-term oral administration of *Amalaki rasayana*, Mol. Cell. Biochem., 463 (2020), pp. 147-160, 10.1007/s11010-019-03637-

[40] R.A. Silva, T.C.S. Pereira, A.R. Souza, P.R. Ribeiro, 1H NMR-based metabolite profiling for biomarker identification, Clin. Chim. Acta, 502 (2020), pp. 269-279, 10.1016/j.cca.2019.11.015

[41] X. Zhang, Q. Li, Z. Xu, J. Dou, Mass spectrometry-based metabolomics in health and medical science: a systematic review, RSC Adv., 10 (2020), pp. 3092-3104, 10.1039/C9RA08985C

[42] F. Fenaille, P. Barbier Saint-Hilaire, K. Rousseau, C. Junot, Data acquisition workflows in liquid chromatography coupled to high resolution mass spectrometry-based metabolomics: where do we stand?, J. Chromatogr. A, 1526 (2017), pp. 1-12, 10.1016/j.chroma.2017.10.043

[43] C. Junot, F. Fenaille, B. Colsch, F. Bécher, High resolution mass spectrometry based techniques at the crossroads of metabolic pathways, Mass Spectrom. Rev., 33 (2014), pp. 471-500, 10.1002/mas.21401

[44] D. Stettin, R.X. Poulin, G. Pohnert, Metabolomics benefits from orbitrap GC–MS—comparison of low- and high-resolution GC–MS, Metabolites, 10 (2020), p. 143, 10.3390/metabo10040143

[45] J. Guo, T. Huan, Comparison of full-scan, data-dependent, and data-independent acquisition modes in liquid chromatography–mass spectrometry based untargeted metabolomics, Anal. Chem., 92 (2020), pp. 8072-8080, 10.1021/acs.analchem.9b05135

[46] D. Licha, S. Vidali, S. Aminzadeh-Gohari, O. Alka, L. Breitkreuz, O. Kohlbacher, R.J. Reischl, R.G. Feichtinger, B. Kofler, C.G. Huber, Untargeted metabolomics reveals molecular effects of ketogenic diet on healthy and tumor *Xenograft* mouse models, Int. J. Mol. Sci., 20 (2019), p. 3873, 10.3390/ijms20163873

[47] J. Tovar, V.D. de Mello, A. Nilsson, M. Johansson, J. Paananen, M. Lehtonen, K. Hanhineva, I. Björck, Reduction in cardiometabolic risk factors by a multifunctional diet is mediated via several branches of metabolism as evidenced by nontargeted metabolite profiling approach, Mol. Nutr. Food Res., 61 (2017), p. 1600552, 10.1002/mnfr.201600552

[48] D.E. Hoyos Ossa, R. Gil-Solsona, G.A. Peñuela, J.V. Sancho, F.J. Hernández, Assessment of protected designation of origin for Colombian coffees based on HRMS-based metabolomics, Food Chem., 250 (2018), pp. 89-97, 10.1016/j.foodchem.2018.01.038

[49] M.R. Showalter, E.B. Nonnecke, A.L. Linderholm, T. Cajka, M.R. Sa, B. Lönnerdal, N.J. Kenyon, O. Fiehn, Obesogenic diets alter metabolism in mice, PloS One, 13 (2018), Article e0190632, 10.1371/journal.pone.0190632

[50] L. Olmo-García, K. Wendt, N. Kessler, A. Bajoub, A. Fernández-Gutiérrez, C. Baessmann, A. Carrasco-Pancorbo, Exploring the capability of LC-MS and GC-MS multi-class methods to discriminate virgin olive oils from different geographical indications and to identify potential origin markers, Eur. J. Lipid Sci. Technol., 121 (2019), 10.1002/ejlt.201800336

[51] M. Khodadadi, M. Pourfarzam, A review of strategies for untargeted urinary metabolomic analysis using gas chromatography–mass spectrometry, Metabolomics, 16 (2020), pp. 1-14, 10.1007/s11306-020-01687-x

[52] D.J. Beale, F.R. Pinu, K.A. Kouremenos, M.M. Poojary, V.K. Narayana, B.A. Boughton, K. Kanojia, S. Dayalan, O.A.H. Jones, D.A. Dias, Review of Recent Developments in GC–MS Approaches to Metabolomics-Based Research, Springer US (2018), 10.1007/s11306-018-1449-2

[53] B.B. Misra, M. Olivier, High resolution GC-orbitrap-MS metabolomics using both electron ionization and chemical ionization for analysis of human plasma, J. Proteome Res. (2020), 10.1021/acs.jproteome.9b00774

[54] P.S.C. Sri Harsha, R. Abdul Wahab, C. Cuparencu, L. Dragsted, L. Brennan, A metabolomics approach to the identification of urinary biomarkers of pea intake, Nutrients, 10 (2018), p. 1911, 10.3390/nu10121911

[55] N. Vázquez-Manjarrez, C.H. Weinert, M.M. Ulaszewska, C.I. Mack, P. Micheau, M. Pétéra, S. Durand, E. Pujos-Guillot, B. Egert, F. Mattivi, A. Bub, L.O. Dragsted, S.E. Kulling, C. Manach, Discovery and validation of banana intake biomarkers using untargeted metabolomics in human intervention and cross-sectional studies, J. Nutr., 149 (2019), pp. 1701-1713, 10.1093/jn/nxz125

[56] G. Rocchetti, L. Lucini, A. Gallo, F. Masoero, M. Trevisan, G. Giuberti, Untargeted metabolomics reveals differences in chemical fingerprints between PDO and non-PDO Grana Padano cheeses, Food Res. Int., 113 (2018), pp. 407-413, 10.1016/j.foodres.2018.07.029

[57] Y. Li, Y. Jin, S. Yang, W. Zhang, J. Zhang, W. Zhao, L. Chen, Y. Wen, Y. Zhang, K. Lu, Y. Zhang, J. Zhou, S. Yang, Strategy for comparative untargeted metabolomics reveals honey markers of different floral and geographic origins using ultrahigh-performance liquid chromatography-hybrid quadrupole-orbitrap mass spectrometry, J. Chromatogr. A, 1499 (2017), pp. 78-89, 10.1016/j.chroma.2017.03.071

[58] H. Ismail, A.L. Gillespie, D. Calderwood, H. Iqbal, C. Gallagher, O.P. Chevallier, C.T. Elliott, X. Pan, B. Mirza, B.D. Green, The health promoting bioactivities of *Lactuca sativa* can be enhanced by genetic modulation of plant secondary metabolites, Metabolites, 9 (2019), p. 97, 10.3390/metabo9050097

[59] M. Wei, Z. Liu, Y. Liu, S. Li, M. Hu, K. Yue, T. Liu, Y. He, Z. Pi, Z. Liu, F. Song, Urinary and plasmatic metabolomics strategy to explore the holistic mechanism of lignans in *S. chinensis* in treating Alzheimer's disease using UPLC-Q-TOF-MS, Food Funct., 10 (2019), pp. 5656-5668, 10.1039/C9FO00677J

[60] H.-A. Oh, H. Lee, S. Park, Y. Lim, O. Kwon, J.Y. Kim, D. Kim, B.H. Jung, Analysis of plasma metabolic profiling and evaluation of the effect

of the intake of *Angelica keiskei* using metabolomics and lipidomics, J. Ethnopharmacol., 243 (2019), p. 112058, 10.1016/j.jep.2019.112058

[61] L. Kozlowska, O. Mizera, A. Mroz, An untargeted metabolomics approach to investigate the metabolic effect of beetroot juice supplementation in Fencers—a preliminary study, Metabolites, 10 (2020), p. 100, 10.3390/metabo10030100

[62] D.K. Trivedi, R.K. Iles, Do not just do it, do it right: urinary metabolomics -establishing clinically relevant baselines, Biomed. Chromatogr., 28 (2014), pp. 1491-1501, 10.1002/bmc.3219

[63] E. da Costa, V. Azevedo, T. Melo, A. Rego, D.V. Evtuguin, P. Domingues, R. Calado, R. Pereira, M. Abreu, M. Domingues, High-resolution lipidomics of the early life stages of the red seaweed *Porphyra dioica*, Molecules, 23 (2018), p. 187, 10.3390/molecules23010187

[64] L. Lacalle-Bergeron, T. Portolés, F.J. López, J.V. Sancho, C. Ortega-Azorín, E.M. Asensio, O. Coltell, D. Corella, Ultra-performance liquid chromatography-ion mobility separation-quadruple time-of-flight MS (UHPLC-IMS-QTOF MS) metabolomics for short-term biomarker discovery of orange intake: a randomized, controlled crossover study, Nutrients, 12 (2020), p. 1916, 10.3390/nu12071916

[65] L. Shi, C. Brunius, I. Johansson, I.A. Bergdahl, B. Lindahl, K. Hanhineva, R. Landberg, Plasma metabolites associated with healthy Nordic dietary indexes and risk of type 2 diabetes—a nested case-control study in a Swedish population, Am. J. Clin. Nutr., 108 (2018), pp. 564-575, 10.1093/ajcn/nqy145

[66] S. Heischmann, L.B. Gano, K. Quinn, L.-P. Liang, J. Klepacki, U. Christians, N. Reisdorph, M. Patel, Regulation of kynurenine metabolism by a ketogenic diet, J. Lipid Res., 59 (2018), pp. 958-966, 10.1194/jlr.M079251

[67] R. Pérez-Míguez, M. Castro-Puyana, E. Sánchez-López, M. Plaza, M.L. Marina, Untargeted HILIC-MS-based metabolomics approach to evaluate coffee roasting process: contributing to an integrated metabolomics multiplatform, Molecules, 25 (2020), p. 887, 10.3390/molecules25040887

[68] A.M. De Livera, D.A. Dias, D. De Souza, T. Rupasinghe, J. Pyke, D. Tull, U. Roessner, M. McConville, T.P. Speed, Normalizing and integrating metabolomics data, Anal. Chem., 84 (2012), pp. 10768-10776, 10.1021/ac302748b

[69] D. Dudzik, C. Barbas-Bernardos, A. García, C. Barbas, Quality assurance procedures for mass spectrometry untargeted metabolomics. a review, J. Pharmaceut. Biomed. Anal., 147 (2018), pp. 149-173, 10.1016/j.jpba.2017.07.044

[70] B. Li, J. Tang, Q. Yang, X. Cui, S. Li, S. Chen, Q. Cao, W. Xue, N. Chen, F. Zhu, Performance evaluation and online realization of data-driven normalization methods used in LC/MS based untargeted metabolomics analysis, Sci. Rep., 6 (2016), pp. 1-13, 10.1038/srep38881

[71] E. Gorrochategui, J. Jaumot, S. Lacorte, R. Tauler, Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: overview and workflow, TrAC Trends Anal. Chem., 82 (2016), pp. 425-442, 10.1016/j.trac.2016.07.004

[72] G. Pimentel, K.J. Burton, U. von Ah, U. Bütikofer, F.P. Pralong, N. Vionnet, R. Portmann, G. Vergères, Metabolic footprinting of fermented milk consumption in serum of healthy men, J. Nutr., 148 (2018), pp. 851-860, 10.1093/jn/nxy053

[73] W. Perng, L. Tang, P.X.K. Song, M. Goran, M.M. Tellez Rojo, A. Cantoral, K.E. Peterson, Urate and nonanoate mark the relationship between sugar-sweetened beverage intake and blood pressure in adolescent girls: a metabolomics analysis in the element cohort, Metabolites, 9 (2019), p. 100, 10.3390/metabo9050100

[74] L. Li, S. Ma, D. Wang, L. Chen, X. Wang, Plasma metabolomics analysis of endogenous and exogenous metabolites in the rat after administration of *Lonicerae Japonicae Flos*, Biomed. Chromatogr., 34 (2020), 10.1002/bmc.4773

[75] G. Delaporte, M. Cladière, D. Jouan-Rimbaud Bouveresse, V. Camel, Untargeted food contaminant detection using UHPLC-HRMS

combined with multivariate analysis: feasibility study on tea, Food Chem., 277 (2019), pp. 54-62, 10.1016/j.foodchem.2018.10.089

[76] B. Warth, P. Raffeiner, A. Granados, T. Huan, M. Fang, E.M. Forsberg, H.P. Benton, L. Goetz, C.H. Johnson, G. Siuzdak, Metabolomics reveals that dietary xenoestrogens alter cellular metabolism induced by palbociclib/letrozole combination cancer therapy, Cell Chem. Biol., 25 (2018), pp. 291-300, 10.1016/j.chembiol.2017.12.010

[77] Y. Hövelmann, A. Jagels, R. Schmid, F. Hübner, H.-U. Humpf, Identification of potential human urinary biomarkers for tomato juice intake by mass spectrometry-based metabolomics, Eur. J. Nutr., 59 (2020), pp. 685-697, 10.1007/s00394-019-01935-4

[78] G. Rizzato, E. Scalabrin, M. Radaelli, G. Capodaglio, O. Piccolo,A new exploration of licorice metabolome, Food Chem., 221 (2017), pp. 959-968, 10.1016/j.foodchem.2016.11.068

[79] M.J. Cichon, K.M. Riedl, L. Wan, J.M. Thomas-Ahner, D.M. Francis, S.K. Clinton, S.J. Schwartz, Plasma metabolomics reveals steroidal alkaloids as novel biomarkers of tomato intake in mice, Mol. Nutr. Food Res., 61 (2017), p. 1700241, 10.1002/mnfr.201700241

[80] P. Keski-Rahkonen, M. Kolehmainen, J. Lappi, V. Micard, J. Jokkala, N. Rosa-Sibakov, J. Pihlajamäki, P.V. Kirjavainen, H. Mykkänen, K. Poutanen, M.J. Gunter, A. Scalbert, K. Hanhineva, Decreased plasma serotonin and other metabolite changes in healthy adults after consumption of wholegrain rye: an untargeted metabolomics study, Am. J. Clin. Nutr., 109 (2019), pp. 1630-1639, 10.1093/ajcn/nqy394

[81] G. Peron, D. Santarossa, D. Voinovich, S. Dall'Acqua, S. Sut, Urine metabolomics shows an induction of fatty acids metabolism in healthy adult volunteers after supplementation with green coffee (*Coffea robusta L.*) bean extract, Phytomedicine, 38 (2018), pp. 74-83, 10.1016/j.phymed.2017.11.002

[82] Z. Qin, C. Jia, D. Liao, X. Chen, X. Li, Comparison of serum metabolite changes of radiated mice administered with *Panax quinquefolium* from different cultivation regions using UPLC-Q/TOF-MS based

metabolomic approach, Molecules, 23 (2018), p. 1014, 10.3390/molecules23051014

[83] D. Cavanna, D. Catellani, C. Dall'Asta, M. Suman, Egg product freshness evaluation: a metabolomic approach, J. Mass Spectrom., 53 (2018), pp. 849-861, 10.1002/jms.4256

[84] H. Tsugawa, T. Cajka, T. Kind, Y. Ma, B. Higgins, K. Ikeda, M. Kanazawa, J. VanderGheynst, O. Fiehn, M. Arita, MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis, Nat. Methods, 12 (2015), pp. 523-526, 10.1038/nmeth.3393

[85] M. Nakanishi, A. Matz, C. Klemashevich, D.W. Rosenberg, Dietary walnut supplementation alters mucosal metabolite profiles during DSS-induced colonic ulceration, Nutrients, 11 (2019), p. 1118, 10.3390/nu11051118

[86] A.M. Gallego, L.F. Rojas, H.A. Rodriguez, C. Mora, L. Atehortúa, A.I. Urrea, M.J. Guiltinan, S.N. Maximova, E. Gaquerel, M. Zuluaga, N. Pabón-Mora, Metabolomic profile of cacao cell suspensions growing in blue light/dark conditions with potential in food biotechnology, Plant Cell Tissue Organ Cult., 139 (2019), pp. 275-294, 10.1007/s11240-019-01679-3

[87] L.G. Johnsen, P.B. Skou, B. Khakimov, R. Bro, Gas chromatography – mass spectrometry data processing made easy, J. Chromatogr. A, 1503 (2017), pp. 57-64, 10.1016/j.chroma.2017.04.052

[88] T. Skov, A.H. Honoré, H.M. Jensen, T. Næs, S.B. Engelsen, Chemometrics in foodomics: handling data structures from multiple analytical platforms, TrAC Trends Anal. Chem., 60 (2014), pp. 71-79, 10.1016/j.trac.2014.05.004

[89] Z. Jandrić, M. Islam, D.K. Singh, A. Cannavan, Authentication of Indian citrus fruit/fruit juices by untargeted and targeted metabolomics, Food Contr., 72 (2017), pp. 181-188, 10.1016/j.foodcont.2015.10.044

[90] D. González-Peña, D. Dudzik, A. García, B. Ancos, C. Barbas, C. Sánchez-Moreno, Metabolomic fingerprinting in the comprehensive study

of liver changes associated with onion supplementation in hypercholesterolemic Wistar rats, Int. J. Mol. Sci., 18 (2017), p. 267, 10.3390/ijms18020267

[91] J. Rubert, A. Monforte, K. Hurkova, G. Pérez-Martínez, J. Blesa, J.L. Navarro, M. Stranka, J.M. Soriano, J. Hajslova, Untargeted metabolomics of fresh and heat treatment Tiger nut (*Cyperus esculentus L.*) milks reveals further insight into food quality and nutrition, J. Chromatogr. A, 1514 (2017), pp. 80-87, 10.1016/j.chroma.2017.07.071

[92] Z. Mi, L. Kwok, J. Xue, Y. Wang, H. Zhang, Y. Chen, Fermentation dynamics of *Lactobacillus helveticus H9* revealed by ultra-performance liquid chromatography quadrupole time-of-flight mass spectrometry, Int. J. Food Sci. Technol., 53 (2018), pp. 1442-1451, 10.1111/ijfs.13723

[93] X. Hu, J.D. Chandler, M.L. Orr, L. Hao, K. Liu, K. Uppal, Y.-M. Go, D.P. Jones, Selenium supplementation alters hepatic energy and fatty acid metabolism in mice, J. Nutr., 148 (2018), pp. 675-684, 10.1093/jn/nxy036

[94] K. Hurkova, L. Uttl, J. Rubert, K. Navratilova, V. Kocourek, M. Stranska-Zachariasova, F. Paprstein, J. Hajslova, Cranberries versus lingonberries: a challenging authentication of similar *Vaccinium* fruit, Food Chem., 284 (2019), pp. 162-170, 10.1016/j.foodchem.2019.01.014

[95] T. Skov, S.B. Engelsen, Chemometrics, Mass Spectrometry, and Foodomics, Foodomics, John Wiley & Sons, Inc., Hoboken, NJ, USA (2013), pp. 507-538, 10.1002/9781118537282.ch19

[96] M. Bevilacqua, R. Bro, F. Marini, Å. Rinnan, M.A. Rasmussen, T. Skov, Recent chemometrics advances for foodomics, TrAC Trends Anal. Chem., 96 (2017), pp. 42-51, 10.1016/j.trac.2017.08.011

[97] J. Hao, H. Hu, J. Liu, X. Wang, X. Liu, J. Wang, M. Niu, Y. Zhao, X. Xiao, Integrated metabolomics and network pharmacology study on immunoregulation mechanisms of *Panax ginseng* through macrophages, evidence-based complement, Altern. Med., 2019 (2019), pp. 1-14, 10.1155/2019/3630260

[98] W.J. Nash, W.B. Dunn, From mass to metabolite in human untargeted metabolomics: recent advances in annotation of metabolites applying liquid chromatography-mass spectrometry data, TrAC Trends Anal. Chem., 120 (2019), p. 115324, 10.1016/j.trac.2018.11.022

[99] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, T.W.M. Fan, O. Fiehn, R. Goodacre, J.L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A.N. Lane, J.C. Lindon, P. Marriott, A.W. Nicholls, M.D. Reily, J.J. Thaden, M.R. Viant, Proposed minimum reporting standards for chemical analysis, Metabolomics, 3 (2007), pp. 211-221, 10.1007/s11306-007-0082-2

[100] E.L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H.P. Singer, J. Hollender, Identifying small molecules via high resolution mass spectrometry: communicating confidence, Environ. Sci. Technol., 48 (2014), pp. 2097-2098, 10.1021/es5002105

[101] A. Celma, J.V. Sancho, E.L. Schymanski, D. Fabregat-Safont, M. Ibáñez, J. Goshawk, G. Barknowitz, F. Hernández, L. Bijlsma, Improving target and suspect screening high-resolution mass spectrometry workflows in environmental analysis by ion mobility separation, Environ. Sci. Technol. (2020), 10.1021/acs.est.0c05713

[102] M. Vinaixa, E.L. Schymanski, S. Neumann, M. Navarro, R.M. Salek, O. Yanes, Mass spectral databases for LC/MS- and GC/MS-based metabolomics: state of the field and future prospects, TrAC Trends Anal. Chem., 78 (2016), pp. 23-35, 10.1016/j.trac.2015.09.005

[103] K. Böhme, P. Calo-Mata, J. Barros-Velázquez, I. Ortea, Recent applications of omics-based technologies to main topics in food authentication, TrAC Trends Anal. Chem., 110 (2019), pp. 221-232, 10.1016/j.trac.2018.11.005

[104] L. Brennan, F.B. Hu, Metabolomics-based dietary biomarkers in nutritional epidemiology—current status and future opportunities, Mol. Nutr. Food Res., 63 (2019), pp. 1-5, 10.1002/mnfr.201701064

[105] M. Guasch-Ferré, S.N. Bhupathiraju, F.B. Hu, Use of metabolomics in improving assessment of dietary intake, Clin. Chem., 64 (2018), pp. 82-98, 10.1373/clinchem.2017.272344

[106] C. Collins, A.E. McNamara, L. Brennan, Role of metabolomics in identification of biomarkers related to food intake, Proc. Nutr. Soc., 78 (2019), pp. 189-196, 10.1017/S002966511900048X

[107] V. Özdemir, E. Kolker, Precision Nutrition 4.0: A Big Data and Ethics Foresight Analysis—Convergence of Agrigenomics, Nutrigenomics, Nutriproteomics, and Nutrimetabolomics, A J. Integr. Biol., 20 (2016), pp. 69-75, 10.1089/omi.2015.0193

[108] D. Braconi, G. Bernardini, L. Millucci, A. Santucci, Foodomics for human health: current status and perspectives, Expert Rev. Proteomics, 15 (2018), pp. 153-164, 10.1080/14789450.2018.1421072

[109] G. Mancano, M. Mora-Ortiz, S.P. Claus, Recent developments in nutrimetabolomics: from food characterisation to disease prevention, Curr. Opin. Food Sci., 22 (2018), pp. 145-152, 10.1016/j.cofs.2018.03.012

[110] C. Claassen, J. Kuballa, S. Rohn, Polar lipids in starch-rich commodities to be analyzed with LC-MS-based metabolomics—optimization of ionization parameters and high-throughput extraction protocols, Metabolites, 9 (2019), p. 167, 10.3390/metabo9080167

[111] R. Gil Solsona, C. Boix, M. Ibáñez, J.V. Sancho, The classification of almonds (*Prunus dulcis*) by country and variety using UHPLC-HRMS-based untargeted metabolomics, Food Addit. Contam. Part A, 35 (2018), pp. 395-403, 10.1080/19440049.2017.1416679

[112] A. O'Gorman, L. Brennan, The role of metabolomics in determination of new dietary biomarkers, Proc. Nutr. Soc., 76 (2017), pp. 295-302, 10.1017/S0029665116002974

[113] J.A. Rothwell, P. Keski-Rahkonen, N. Robinot, N. Assi, C. Casagrande, M. Jenab, P. Ferrari, M.C. Boutron-Ruault, Y. Mahamat-Saleh, F.R. Mancini, H. Boeing, V. Katzke, T. Kühn, K. Niforou, A. Trichopoulou, E. Valanou, V. Krogh, A. Mattiello, D. Palli, C. Sacerdote, R. Tumino, A. Scalbert, A metabolomic study of biomarkers of habitual coffee intake in four European

countries, Mol. Nutr. Food Res., 63 (2019), pp. 1-10, 10.1002/mnfr.201900659

[114] E. Acar, G. Gürdeniz, B. Khakimov, F. Savorani, S.K. Korndal, T.M. Larsen, S.B. Engelsen, A. Astrup, L.O. Dragsted, Biomarkers of individual foods, and separation of diets using untargeted LC-MS-based plasma metabolomics in a randomized controlled trial, Mol. Nutr. Food Res., 63 (2019), p. 1800215, 10.1002/mnfr.201800215

[115] L.O. Dragsted, Q. Gao, A. Scalbert, G. Vergères, M. Kolehmainen, C. Manach, L. Brennan, L.A. Afman, D.S. Wishart, C. Andres Lacueva, M. Garcia-Aloy, H. Verhagen, E.J.M. Feskens, G. Praticò, Validation of biomarkers of food intake-Critical assessment of candidate biomarkers, Genes Nutr., 13 (2018), pp. 1-14, 10.1186/s12263-018-0603-9

[116] C. Picó, F. Serra, A.M. Rodríguez, J. Keijer, A. Palou, Biomarkers of nutrition and health: new tools for new approaches, Nutrients, 11 (2019), p. 1092, 10.3390/nu11051092

[117] A. Wells, W.T. Barrington, S. Dearth, A. May, D.W. Threadgill, S.R. Campagna, B.H. Voy, Tissue level diet and sex-by-diet interactions reveal unique metabolite and clustering profiles using untargeted liquid chromatography-mass spectrometry on adipose, skeletal muscle, and liver tissue in C57BL6/J mice, J. Proteome Res., 17 (2018), pp. 1077-1090, 10.1021/acs.jproteome.7b00750

[118] Y.J. Weng, H.Y. Gan, X. Li, Y. Huang, Z.C. Li, H.M. Deng, S.Z. Chen, Y. Zhou, L.S. Wang, Y.P. Han, Y.F. Tan, Y.J. Song, Z.M. Du, Y.Y. Liu, Y. Wang, N. Qin, Y. Bai, R.F. Yang, Y.J. Bi, F.C. Zhi, Correlation of diet, microbiota and metabolite networks in inflammatory bowel disease, J. Dig. Dis., 20 (2019), pp. 447-459, 10.1111/1751-2980.12795

[119] L. Shi, C. Brunius, I.A. Bergdahl, I. Johansson, O. Rolandsson, C. Donat Vargas, H. Kiviranta, K. Hanhineva, A. Åkesson, R. Landberg, Joint analysis of metabolite markers of fish intake and persistent organic pollutants in relation to type 2 diabetes risk in Swedish adults, J. Nutr., 149 (2019), pp. 1413-1423, 10.1093/jn/nxz068

[120] L. Shi, C. Brunius, I. Johansson, I.A. Bergdahl, O. Rolandsson, B. Guelpen, A. Winkvist, K. Hanhineva, R. Landberg, Plasma metabolite biomarkers of boiled and filtered coffee intake and their association with type 2 diabetes risk, J. Intern. Med., 287 (2020), pp. 405-421, 10.1111/joim.13009

[121] L. Wang, D. Yao, P.E. Urriola, A.R. Hanson, M. Saqui-Salces, B.J. Kerr, G.C. Shurson, C. Chen, Identification of activation of tryptophan–NAD+ pathway as a prominent metabolic response to thermally oxidized oil through metabolomics-guided biochemical analysis, J. Nutr. Biochem., 57 (2018), pp. 255-267, 10.1016/j.jnutbio.2018.04.009

[122] S.P. Claus, Development of personalized functional foods needs metabolic profiling, Curr. Opin. Clin. Nutr. Metab. Care, 17 (2014), pp. 567-573, 10.1097/MCO.0000000000000107

[123] A. Valdés, A. Cifuentes, C. León, Foodomics evaluation of bioactive compounds in foods, TrAC Trends Anal. Chem., 96 (2017), pp. 2-13, 10.1016/j.trac.2017.06.004

[124] D. Rago, M.A. Rasmussen, K.A. Lee-Sarwar, S.T. Weiss, J. Lasky-Su, J. Stokholm, K. Bønnelykke, B.L. Chawes, H. Bisgaard, Fish-oil supplementation in pregnancy, child metabolomics and asthma risk, EBioMedicine, 46 (2019), pp. 399-410, 10.1016/j.ebiom.2019.07.057

[125] The ultimate workflow for small molecule discovery, 66 (2020), https://www.thermofisher.com/es/es/home/global/forms/industrial/ultimate-workflow-small-molecule-discovery.html

[126] F. Hernández, M. Ibáñez, T. Portolés, M.I. Cervera, J.V. Sancho, F.J. López, Advancing towards universal screening for organic pollutants in waters, J. Hazard Mater., 282 (2015), pp. 86-95, 10.1016/j.jhazmat.2014.08.006

1.3 References

Bowen, B. P., & Northen, T. R. (2010). Dealing with the unknown: Metabolomics and Metabolite Atlases. Journal of the American Society for Mass Spectrometry, 21(9), 1471–1476. https://doi.org/10.1016/j.jasms.2010.04.003

Cifuentes, A. (2013). Foodomics: principles and applications. In A. Cifuentes (Ed.), Foodomics: advanced mass spectrometry in modern food sciences and nutrition (p. 580). John Wiley & Sons, Inc. https://www.wiley.com/en-us/Foodomics%3A+Advanced+Mass+Spectrometry+in+Modern+Food+Science+and+Nutrition-p-9781118169452

Dettmer, K., Aronov, P. A., & Hammock, B. D. (2007). Mass spectrometry-based metabolomics. Mass Spectrometry Reviews, 26(1), 51–78. https://doi.org/10.1002/mas.20108

Dunn, W. B., & Ellis, D. I. (2005). Metabolomics: Current analytical platforms and methodologies. TrAC Trends in Analytical Chemistry, 24(4), 285–294. https://doi.org/10.1016/j.trac.2004.11.021

Fiehn, O. (2002). Metabolomics - The link between genotypes and phenotypes. Plant Molecular Biology, 48(1–2), 155–171. https://doi.org/10.1023/A:1013713905833

Karaman, I. (2017). Preprocessing and Pretreatment of Metabolomics Data for Statistical Analysis. In Advances in Experimental Medicine and Biology (Vol. 965, pp. 145–161). Springer New York LLC. https://doi.org/10.1007/978-3-319-47656-8_6

Kuehnbaum, N. L., & Britz-McKibbin, P. (2013). New advances in separation science for metabolomics: resolving chemical diversity in a post-genomic era. Chemical Reviews, 113(4), 2437–2468. https://doi.org/10.1021/cr300484s

Lindon, J. C., Holmes, E., & Nicholson, J. K. (2003). Peer Reviewed: So What's the Deal with Metabonomics? Analytical Chemistry, 75(17), 384 A-391 A. https://doi.org/10.1021/ac031386+

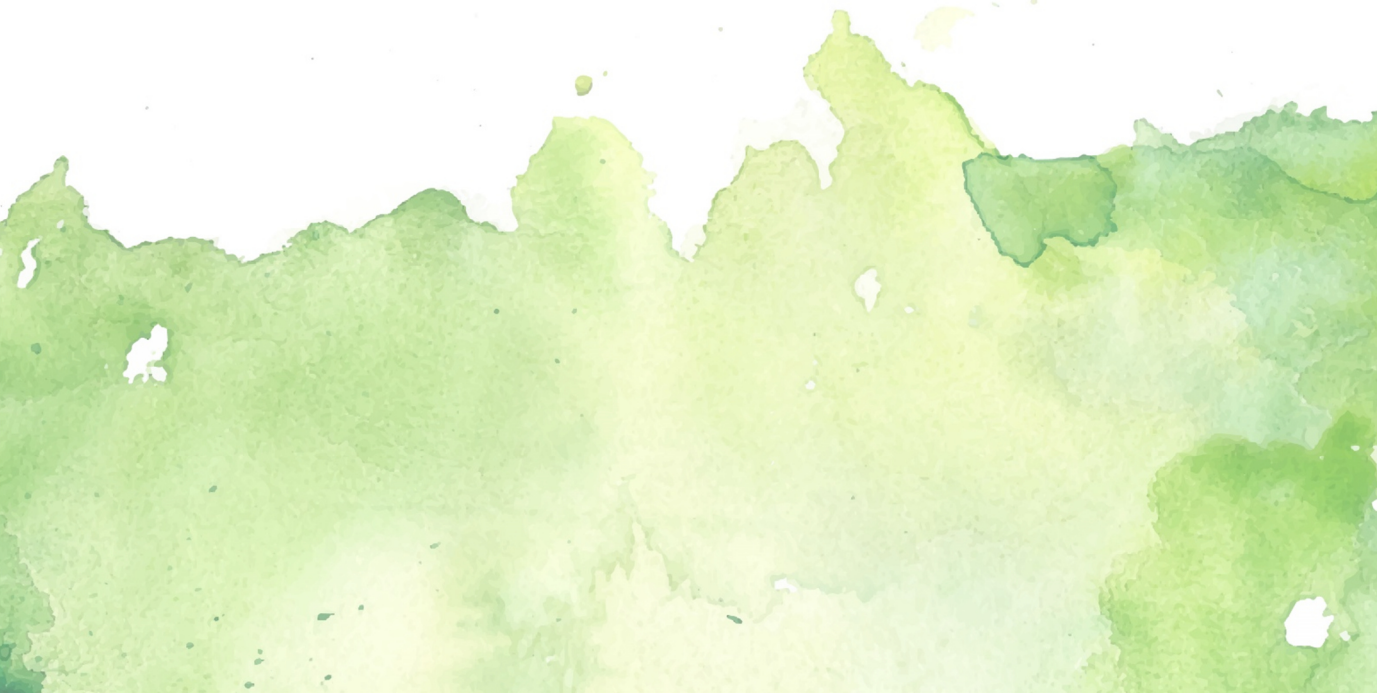
Oliver, S. G., Winson, M. K., Kell, D. B., & Baganz, F. (1998). Systematic functional analysis of the yeast genome. Trends in Biotechnology, 16(9), 373–378. https://doi.org/10.1016/s0167-7799(98)01214-1

Pinto, R. C. (2017). Chemometrics methods and strategies in metabolomics. In Advances in Experimental Medicine and Biology (Vol. 965, pp. 163–190). Springer New York LLC. https://doi.org/10.1007/978-3-319-47656-8_7

Veenstra, T. D. (2021). Omics in Systems Biology: Current Progress and Future Outlook. PROTEOMICS, 21(3–4), 2000235. https://doi.org/10.1002/pmic.202000235

Weckwerth, W. (2003). Metabolomics in Sytems Biology. Annual Review of Plant Biology, 54(1), 669–689. https://doi.org/10.1146/annurev.arplant.54.031902.135014

Chapter 2.
Food processing



Chapter 2. Food Processing

2.1 Introduction

2.2 Scientific article II

"Gas chromatography-mass spectrometry based untargeted volatolomics for smoked seafood classification"

Food Research International 137 (2020) 109698

2.3 Discussion of the results

2.4 References

2.1 Introduction

Modern food industry have to face many new challenges every day related to the demand of safety control, quality assesment and traceability of food products from consumers, and consequently from the authorities. In response, both academia and food industries have implement many powerful analytical strategies for the determination of biomarkers of food origin and authenticity to guarantee the confidence and well-being of consumers; and for the early warning of food safety issues that may jeopardize consumers' health.

Some of these problems have been traditionally addressed in a targeted way, such as the quantification or screening of potentially dangerous compounds, like environmental pollutants, pesticides residues, veterinary drugs, toxicants, allergens among others (Castro-Puyana et al., 2017). The regulations that established around these compounds and their concentration limits in food are constantly updated, evolving along with science and new discoveries.

However, untargeted metabolomics is a rising approach in the different areas related to food processing (Jacobs et al., 2021). Due to the high diversity of samples matrices (foods) and variety of parameters for the assessment of quality, traceability and safety, the global screening approximation of untargeted metabolomics is gaining interest, especially its combination with chromatographic techniques coupled to mass spectrometry for the analysis.

Untargeted metabolomics has been applied to assess food quality related parameters as food composition (Lozano-Castellón et al., 2021), aroma (Nyarukowa et al., 2021) and taste (Li et al., 2021). These parameters have a huge variation from a product to another, they are not completely objectives and they are dependent on multiple considerations about the product itself and even the consumer preferences. These evaluations grant higher value in the market, hence the interest in ensuring the quality of food products. Untargeted metabolomics has also been appplied to determine other parameters highly related with food quality as phytochemical composition with benefical effects on healt of fresh fruit juices in comparison to pasterurised ones (Wang & Xu, 2022); the selection of post-harvest

processing (Qiu et al., 2021); or the effects of hens dietary supplemetation onto the eggs (Giannenas et al., 2021).

Some of the considerations to qualify a product are interrelated with the origin designation (i.e. considerating a variety from one place better than others), authenticity (being a product more valuated than another) and obviously the safety. Rivera-Pérez et al found origin discriminant markers of black pepper (Rivera-Pérez et al., 2021). Brigante et al. compared the fingerprint of chia, flax and sesame seeds for the autetification of bakery products with the added value of these ingredients, known for their nutritional value (Brigante et al., 2022). The impact on quality and safety about effectiveness conservation of packaging for pre-cooked chicken filets was evaluated by Rangel-Huerta et al (Rangel-Huerta et al., 2022).

In the study included in this chapter, untargeted metabolomics applied to the volatile composition (untargeted volatolomics) from different smoked fishery products allowed the building of a classification model to differentiate between smoking processes (**Scientific Article II**). The model developed aims to differentiate between fish products with different degrees of smoking by the cold smoke procedure, versus treatments with carbon monoxide (CO) and filtered wood smoke. This model is not only relevant in terms of authenticity of cold smoked fish products, but also it has relevance in term of safety as CO and filtered wood are not permitted in the European Union to be considered risky.

2.2 Scientific article II

Food Research International 137 (2020) 109698

Contents lists available at ScienceDirect

Food Research International

journal homepage: www.elsevier.com/locate/foodres

Gas chromatography-mass spectrometry based untargeted volatolomics for smoked seafood classification

Leticia Lacalle-Bergeron [a], Tania Portolés [a,*], Carlos Sales [a], M. Carmen Corell [b], Fernando Domínguez [c], Joaquim Beltrán [a], Juan Vicente Sancho [a], Félix Hernández [a]

[a] Environmental and Public Health Analytical Chemistry, Research Institute for Pesticides and Water (IUPA), University Jaume I, Av. Sos Baynat S/N, 12071 Castellón de la Plana, Spain
[b] Sea Delight Europe, S.L, C/ Sao Paulo, 14 Planta 2ª Oficina nº3-P.I. El Sebadal, 35008 Las Palmas de Gran Canaria, Spain
[c] Food Safety Unit, Centre of Public Health of Castellón, Av. del Mar 12, 12003 Castellón de la Plana, Spain

ARTICLE INFO

Keywords:
Dynamic headspace
Smoked fish
GC/MS
PARAFAC
Volatolomics

ABSTRACT

With the increase of the demand of low flavouring smoked seafood products, there is a need of methodologies able to distinguish between different seafood treatments, as not all of them are allowed in all markers.

Following this objective, in the present work an untargeted volatolomics approach was applied to identify volatile markers that demonstrate that Cold smoked products can be distinguished from Tasteless smoke neither Carbon monoxide treated seafood, which are prohibited in the European Union.

The use of dynamic headspace for the volatile extraction followed by thermal desorption in combination with Gas Chromatography (GC) coupled to single quadrupole Mass Spectrometry (MS) has been employed for the determination of volatile composition of smoked fish. Data processing consisted on the use of PARADISe software, applied for GC/MS data treatment, followed by the multivariate analysis with PLS_Toolbox (MATLAB), and finally the creation and validation of statistical classification model.

All 107 variables obtained allowed the construction of a model reaching the correct classification of 97% of the blind samples, while a simplified model with only 11 variables correctly classified up to 93% of the blind samples. These 11 compounds were elucidated to develop subsequent target volatolomics approaches, if needed. Ordered according to the importance in the classification model, the elucidated compounds were: 3-methylcyclopentanone, ethylbenzene, 2-methyl-2-cyclopenten-1-one, 2-methyl-benzofuran, furfuryl alcohol, 2-acetylfuran, acetophenone, guaiacol, 1-hydroxy-2-butanone, 4-vinylguaicol and acetoin.

The results demonstrated the great potential of untargeted volatolomics for smoked seafood treatments classification.

1. Introduction

The smoking treatment of food products has been applied for food preservation since ancient times. This technique allows the preservation of fish by drying and by adding naturally produced microbistatic constituents from the wood smoke. Nowadays, the smoking techniques have been evolving and the aim of smoking, in addition to preservation, is to develop particular flavour, colour and texture characteristics derived from the burned wood that is used in the process (Varlet, Serot, & Prost, 2009). To this aim, different parameters, such as the type of wood and time of exposure to smoke, among others, have to be optimized by the industry to obtain a certain type of flavour, flavour intensity and product quality (Jónsdóttir, Ólafsdóttir, Chanie, & Haugen, 2008).

Alternative processes have emerged, such as the treatment with carbon monoxide (CO) or filtered wood smoke (tasteless smoke (TS), clear smoke…). These techniques are based on the use of CO as colour-stabilizer, maintaining and enhancing the red colour associated with a fresh aspect of the fish flesh, particularly in tuna, and delaying the browning that usually appears with product aging (Barstad, Alvik, & Løvaas, 2006; Bartolucci et al., 2010; Gokoglu, 2020). Nevertheless, filtered wood smoke and CO-treatments of fish are not permitted in the European Union; moreover, CO is excluded from the list of allowed food additives (Regulation (EC) No 1333/2008). This measure was taken because consumers could be confused about the freshness of the product

* Corresponding author.
E-mail address: tportole@uji.es (T. Portolés).

https://doi.org/10.1016/j.foodres.2020.109698
Received 11 May 2020; Received in revised form 6 August 2020; Accepted 6 September 2020
Available online 18 September 2020

Food Research International 137 (2020) 109698

Gas chromatography-mass spectrometry based untargeted volatolomics for smoked seafood classification

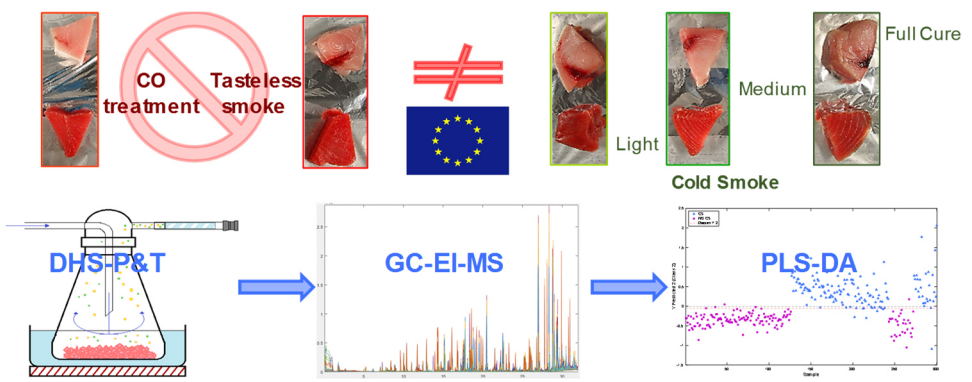
Leticia Lacalle-Bergeron[a], Tania Portolés[a]*, Carlos Sales[a], M.Carmen Corell[b], Fernando Domínguez[c], Joaquim Beltrán[a], Juan V. Sancho[a], Félix Hernández[a]

[a] Enviromental and Public Health Analytical Chemistry, Research Institute for Pesticides and Water (IUPA), University Jaume I, Av. Sos Baynat S/N, 12071 Castellón de la Plana, Spain

[b] Sea Delight Europe, S.L, C/ Sao Paulo, 14 Planta 2ª Oficina nº3-P.I. El Sebadal, 35008 Las Palmas de Gran Canaria, Spain

[c] Food Safety Unit, Centre of Public Health of Castellón, Av. del Mar 12, 12003 Castellón de la Plana, Spain

Graphical abstract



Highlights

- Volatolomic-based approach as powerful methodology for smoked products classification.
- Dynamic headspace-GC/MS is an efficient procedure for volatile compound analysis.
- Volatile profile enabled the correct classification of more than 90 % of the samples.

Abstract

With the increase of the demand of low flavouring smoked seafood products, there is a need of methodologies able to distinguish between different seafood treatments, as not all of them are allowed in all markers. Following this objective, in the present work an untargeted volatolomics approach was applied to identify volatile markers that demonstrate that Cold smoked products can be distinguished from Tasteless smoke neither Carbon monoxide treated seafood, which are prohibited in the European Union. The use of dynamic headspace for the volatile extraction followed by thermal desorption in combination with Gas Chromatography (GC) coupled to single quadrupole Mass Spectrometry (MS) has been employed for the determination of volatile composition of smoked fish. Data processing consisted on the use of PARADISe software, applied for GC/MS data treatment, followed by the multivariate analysis with PLS_Toolbox (MATLAB), and finally the creation and validation of statistical classification model. All 107 variables obtained allowed the construction of a model reaching the correct classification of 97 % of the blind samples, while a simplified model with only 11 variables correctly classified up to 93 % of the blind samples. These 11 compounds were elucidated to develop subsequent target volatolomics approaches, if needed. Ordered according to the importance in the classification model, the elucidated compounds were: 3-methyl-cyclopentanone, ethylbenzene, 2-methyl-2-cyclopenten-1-one, 2-methyl-benzofuran, furfuryl alcohol, 2-acetylfuran, acetophenone, guaiacol, 1-hydroxy-2-butanone, 4-vinylguaiacol and acetoin. The results demonstrated the great potential of untargeted volatolomics for smoked seafood treatments classification.

Keywords

Dynamic headspace; Smoked fish; GC/MS; PARAFAC; Volatolomics

1. Introduction

The smoking treatment of food products has been applied for food preservation since ancient times. This technique allows the preservation of fish by drying and by adding naturally produced microbistatic constituents from the wood smoke. Nowadays, the smoking techniques have been evolving and the aim of smoking, in addition to preservation, is to develop particular flavour, colour and texture characteristics derived from the burned wood that is used in the process (Varlet, Serot, & Prost, 2009). To this aim, different parameters, such as the type of wood and time of exposure to smoke, among others, have to be optimised by the industry to obtain a certain type of flavour, flavour intensity and product quality (Jónsdóttir, Ólafsdóttir, Chanie, & Haugen, 2008).

Alternative processes have emerged, such as the treatment with carbon monoxide (CO) or filtered wood smoke (tasteless smoke (TS), clear smoke…). These techniques are based on the use of CO as colour-stabilizer, maintaining and enhancing the red colour associated with a fresh aspect of the fish flesh, particularly in tuna, and delaying the browning that usually appears with product aging (Barstad et al., 2006, Bartolucci et al., 2010, Gokoglu, 2020). Nevertheless, filtered wood smoke and CO-treatments of fish are not permitted in the European Union; moreover, CO is excluded from the list of allowed food additives (Regulation (EC) No 1333/2008). This measure was taken because consumers could be confused about the freshness of the product (Directive 91/493/EEC). In the case of histidine-rich fishes, the fraudulent use of these treatments may increase the risk of histamine intoxication (Bartolucci et al., 2010, Dalgaard et al., 2008).

The increase in the fish demanded for preparation of sushi in EU has favoured new cold smoking treatments and the application of different degrees of smoking to get different flavours levels. However, because of the faint organoleptic properties of some of them, they can be mistaken by CO or TS smoked products and therefore be rejected at European customs.

The chemicals responsible for the sensory attributes of smoked fish products are mainly volatile organic compounds (VOCs), such as phenols, furan-like compounds, aldehydes or ketones, among others (Varlet et al.,

2009). Accordingly, the analysis of the smoked chemical profile is commonly made by gas chromatography (GC), which in combination with mass spectrometry (GC/MS) allows a sensitive determination with great identification capability able to detect and identify the volatile compounds that characterize the flavour of studied samples.

The extraction technique is a matter of concern, as this step is essential to obtain reliable data and full characterization of the volatile profile. Automatable direct headspace injection (HS) based techniques are commonly used as it implies low sample manipulation, simple, cheap and fast option. Although it suffers from low sensitivity for some compounds present at low concentrations in the vapour phase and the analysis parameters should be carefully optimized to get reproducible results (Soria, García-Sarrió, Ruiz-Matute, & Sanz, 2017). Headspace-solid phase microextraction (HS-SPME) has been already used for VOCs analysis of smoked food products (Marušić Radovčić et al., 2016, Saldaña et al., 2019, Vidal et al., 2017) with a good pre-concentration factor and solventless. In contrast, the adsorption capability is highly dependent on sample matrix and the coating of the fibre (Płotka-Wasylka, Szczepańska, Owczarek, & Namieśnik, 2017). On the contrary, dynamic HS with sorbent trapping (DHS-P&T) captures the VOCs present in the sample on a solid sorbent with the aid of an inert gas flow for continuous extraction. DHS-P&T allows to increase the volatiles recovery, pre-concentrating most of them and therefore enhancing the sensitivity, with good efficiency and low sample manipulation (Soria et al., 2017, Thomsen et al., 2016). Analytes are transferred to the GC system via thermal desorption and then cryo-focused into the GC injector, which leads to an additional increase of the sensitivity due to the complete transfer of the extracted analytes. This technique has been successfully applied to analyse VOCs in food matrices including smoked food (Dirinck et al., 1977, Fredes et al., 2016, Huang et al., 2019, Sales et al., 2019; Soria et al., 2008, Thomsen et al., 2016).

The determination of the volatile profile has been widely applied for food characterization (Beaulieu and Lea, 2006, Ben Brahim et al., 2018, Jónsdóttir et al., 2008, Sérot et al., 2004). Commonly, a targeted approach is used (i.e. focusing the (quantitative) analysis on a limited list of target compounds), which can provide biased information because the compounds

of interest have to be pre-selected and it provides incomplete information on the samples composition. Oppositely to target approaches, the non-targeted metabolomics have great potential in the volatile fingerprinting determination. The metabolomic fingerprinting is defined as "*the unbiased, global screening approach to classify samples based on metabolite patterns or "fingerprints" that change in response to disease, environmental or genetic perturbations with the ultimate goal to identify discriminating metabolites*" (Dettmer, Aronov, & Hammock, 2007). When the object of study is the highly and semi volatile fraction of molecules, it is called volatolomics. This approach was firstly implemented in the health area (Bouhlel et al., 2017, Broza et al., 2015), but it is gaining relevance in food related areas, such as food quality control or food authenticity (Abou-el-karam et al., 2017, Sales et al., 2019). In untargeted volatolomics, the chromatographic analysis must be robust, with adequate peak resolution, and good retention time and peak shape reproducibility. To obtain overall information from the analytes present in samples, and to identify those that might be useful markers for metabolomics, the MS acquisition must be performed in full scan mode (Garcia & Barbas, 2011).

Data processing is of special relevance in non-targeted metabolomics/volatolomics, as reflected in the amount of papers and software reported on metabolomics studies. Some of these software are MzMine (Sales et al., 2017), XCMS (Gil-Solsona et al., 2016), MetAlign (Tomita, Nakamura, & Okada, 2018) or ADMIS (Dudzik et al., 2017) among others. Using these software tools the main objective is to detect the thousands of signals through the chromatogram and obtaining chromatographic peaks at different *m/z*. The peak picking and deconvolution allow to detect relevant *m/z* values associated with a specific retention time, i.e. with a specific component, with a minimum area or intensity. When all samples have been submitted to peak picking, a retention time alignment is performed to match the peaks across the samples (Dudzik, Barbas-Bernardos, García, & Barbas, 2018).

PARAllel FACtor Analysis2 (PARAFAC2) (Johnsen, Amigo, Skov, & Bro, 2014) based Deconvolution and Identification System (PARADISe) (Johnsen, Skou, Khakimov, & Bro, 2017) have appeared recently as new and

innovative application for GC-EI-MS data processing. Differently to other tools, such as XCMS or MzMine, PARADISe performs automatic tentative peak identification based on deconvoluted EI mass spectra in combination with the NIST library. Therefore, it reduces the data matrix as well as the time consumed in statistical analysis and elucidation steps.

In the present study, a volatolomics approach based on GC/MS analysis has been applied to develop a classification model to differentiate between fish product samples (tuna and swordfish) that were submitted to different smoking processes, with modifications in the type of treatment and its intensity. An untargeted volatolomics approach has been applied in this work, contrarily to the works reported until now on VOCs in smoked fish that were based on targeted strategies.

2. Materials and methods

2.1. Chemicals and reagents

Internal standard toluene-D8 (≥99 %) was purchased from Sigma Aldrich (Madrid, Spain). Hexane (trace Analysis quality (AT) GC) was provided by Scharlau (Barcelona, Spain). External standards of volatile compounds, supplied by Supelco (Sigma-Aldrich and Fluka; Barcelona, Spain) as pure compound (92 – 99.5 %) were applied for signal deviation correction purposes and identity confirmation: Toluene, hexanal, butyl acetate, trans-2-hexen-1-al, benzaldehyde, 3-carene, β-cyclocitral, eugenol, β-ionone, 2-isobutylthiazole, (Z)-2-heptenal, α-terpineol, ethylbenzene, (+/−)-3-methylcyclopentanone, acetoin (3-hydroxy-2-butanone) , 1-hydroxy-2-butanone, 2-furyl methyl ketone (2-acetylfuran), 2-methylbenzofuran, furfuryl alcohol (2-furanmethanol), 2-methyl-2-cyclopenten-1-one, acetophenone, 2-methoxy-phenol (guaiacol) and 2-methoxy-4-vinylphenol (4-vinyl guaiacol). Tenax® TA TDU glass thermal desorption tubes 60/80 mesh, O.D. 6 mm × 4 mm (i.d.) × L 60 mm, used as traps were purchased from Gerstel (Mülheim an der Ruhr, Germany).

2.2. Smoked fish samples

The smoked fish samples were provided by Sea Delight Europe, SL. This company has patented a new method of cold smoked where the temperature of the process is maintained at 4 °C with the objective of avoiding

the generation of histamine in blue fish, which appears at temperatures above 4.4 °C. With this method, three types of Cold Smoked seafood products are produced: (i) light smoke grade: the time of flavoured wood smoke exposition is short and the smoked flavour obtained is very light; (ii) medium smoke grade: the exposition to the flavoured wood smoke is larger and the flavour obtained is moderate; (iii) full cure (classic) smoked grade: the product has strong flavour and aroma due to the 48 h curation. Sea Delight produces CO and Tasteless smoked seafood products that are commercialised in USA and Canada, respectively.

A total of 300 samples were used: 20 samples of tuna and 20 samples of swordfish for the Light Cold Smoke (LCS) treatment, and 26 tuna and 26 swordfish samples for each of the other smoking treatments: Tasteless smoke (TS), Carbon monoxide smoke (CO), Full Cure Cold Smoke (FCS), Medium Cold Smoke (MCS) and raw samples (no smoking treatment) (NAT). Samples were stored in freezer at −25 °C until the extraction.

2.3. Purge-and-trap extraction

Smoked fish samples were defrosted at room temperature (24 °C) and triturated before extraction. Then, 5 g of sample were weighed into a 150 mL conical flask. The volatile's extraction procedure was based on our previous works (Beltran et al., 2006, Sales et al., 2019). The flask was immediately closed with a glass tap with two connection tubes: one for the dry N_2 gas entrance and the other for the exit connected to the sorbent Tenax® TA TDU trap tubes (**Fig. S1**). The sorbent trap tubes were previously spiked with 10 μL of 50 μg mL^{-1} internal standard to correct for potential extraction deviations. Sample extraction was carried out at 40 °C for 60 min (immersed in a water bath) with a dry nitrogen (99.7 %) flow of 100 mL min^{-1} to perform the purge process. Finally, the sorbent trap tubes were thermally desorbed with the aid of a TDU into the GC/MS.

The samples were randomly analysed in order to avoid bias in the methodology, performing 18 extractions per day of the samples defrosted immediately before extraction. Quality Control (QCs) samples commonly used in metabolomics (i.e. a pool of samples to monitor the performance of the metabolomic workflow) could not be performed due to the difficulty to

achieve an average representative and homogeneous mix of samples, and to the absence of sample extracts because extraction was made directly in phase gas.

Alternatively, replicate thermal desorption traps were spiked with 10 µL of a mix of volatile compounds at 50 µg mL^{-1}, which were processed at the beginning and at the end of the sequence batch, and every 6 samples, for correction of the instrument deviation.

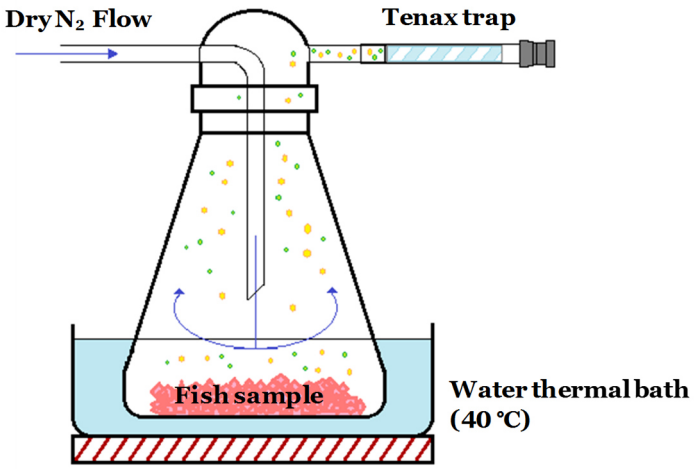


Fig. S1. *Experimental design used for DHS-P&T extraction for smoked fish samples.*

2.4. GC-EI-MS analysis

An Agilent 6890 Plus Series gas chromatograph coupled to a quadrupole mass spectrometer, Agilent 5973 N Mass Selective Detector, with an electron ionization (EI) source and MPS2 autosampler from Gerstel (Linthicum, MD, USA) was used for VOCs analysis. The GC separation was carried out on a 30 m × 0.25 mm DB-WAXETR (0.25 µm film thickness) capillary column (J&W Scientific, Folsom, CA, USA), with helium at a constant flow of 1 mL min^{-1} as carrier gas. The column temperature program started at 40 °C for 3 min; then increased to 160 °C at 5 °C min^{-1} and held for 2 min; then increased to 260 °C at 40 °C min^{-1} and held for 1.50 min (total chromatographic run 32 min).

The injection system comprised two devices; a thermal desorption unit (TDU) and CIS 4 PTV injector. The sorbent traps, used for sample extraction, were thermally desorbed in the TDU in splitless mode using a desorption

program that started at 50 °C (1 min equilibrium time), then increased to 260 °C at 12 °C s^{-1} and held for 8 min; the transfer line temperature was 260 °C. The CIS4 PTV was equipped with a Tenax® TA packed liner and temperature program started at 40 °C during 1 min and then the temperature increased at 12 °C s^{-1} to 260 °C and held for 8 min. In **Fig. S2** a total ion chromatogram obtained from a LCS tuna sample after the described method is shown. As it can be observed, proper chromatographic peak shape is obtained even for the early eluting peaks, although no crio-focusing was used by means of a gap column.

2.5. Data treatment

The GC/MS data, acquired in full SCAN mode, were converted to *.cdf* data format thanks to the Chemstation® (Agilent) "*export to .AIA"* function and pre-processed using the PARADISe software. After loading the exported data to PARADISe software, around 150 time intervals or regions of interest (ROIs) were defined manually along the chromatogram, taking into account the peak shape in total ion chromatogram, when visible, and avoiding empty spaces between intervals.

For each ROI, software calculates a model with a maximum of 8 components and 50,000 maximum iterations in order to resolve the underlying and, possibly, overlapping compounds. Once the model for each interval was created it was optimised with the selection of as many compounds as possible, providing a model fitting and model consistency over 95 % as well as background removal and avoiding model overfitting. The final report was created with the list of compounds and their peak area for each sample in *.xls* format. Since a mixture of external standards were injected every 6 samples, the peak areas were normalised with the area of the closest compound in retention time of the nearest external standard mixture injection, to correct the differences due to instrumental drift, to finally be scaled applying pareto-scaling (van den Berg, Hoefsloot, Westerhuis, Smilde, & van der Werf, 2006).

The statistical multivariate analysis was performed with MATLAB environment (version R2013a, The Mathworks, Natick, MA) along with PLS_Toolbox (version 7.5.2, Eigenvector Research, Wenatchee, WA).

3. Results and discussion

3.1. Volatile extraction procedure performance

Both, direct HS injection and DHS-P&T, were tested for extraction of VOCs. Both systems had been previously used in our laboratory for extraction of volatile compounds in food or food products (Beltran et al., 2006, Fredes et al., 2016, Sales et al., 2019). Although HS-SPME has demonstrated to be a valuable technique for volatile extraction, automatization equipment connected to the GC/MS was not available in our laboratory. Therefore, direct HS injection and DHS-P&T were assayed analysing three replicates of each type of smoked tuna (LCS, MCS, FCS, TS and CO) and raw tuna (NAT) under the same conditions, performing the subsequent analysis by GC/MS in full scan mode.

The results were clearly better (both in terms of sensitivity and number of detected compounds) with the DHS-P&T. The performance obtained with HS static procedure was poorer as regards the number of peaks and the sensitivity reached, which was far below the DHS-P&T which sensitivity was favoured by the pre-concentration factor of the dynamic process, especially for those compounds with low vapour pressure. This fact is in accordance with previous studies (Beltran et al., 2006, Fredes et al., 2016), where a more exhaustive comparison of the available sample treatment for VOCs extraction were performed, as in Sales et al., 2019 (Sales et al., 2019) where the extraction performance of DHS-P&T was tested by obtaining up to 1000 times more sensitivity for most of volatile components compared with a static technique headspace-stir bar sorptive extraction (SHS-SBSE). All together demonstrates the advantages of DHS-P&T over the static headspace techniques.

3.2. Processing of GC/MS data with PARADISe software

Data processing started with GC/MS data conversion to *.cdf* (*netCDF*) format using Chemsation® by Agilent Technologies. Due to its potential in gas chromatography applications and based on our previous experience in our laboratory, PARADISe (Johnsen et al., 2014, Johnsen et al., 2017, Khakimov et al., 2016) was used for peak picking and subsequent alignment of retention times to match the peaks across samples.

To this purpose, the chromatogram was first divided into 150 ROIs and each ROI was individually modelled using the PARAFAC2 algorithm (Harshman, 1972) for peak deconvolution based on the mass spectra and the intensity of the signals. The model validation was conducted following Khakimov et al. (Khakimov et al., 2016) recommendations. For each ROI, the model fitting was tested to a maximum of eight components, selecting the optimal number of components based on a good model fit and core consistency (both over 95 %), noise removal and low residuals, but avoiding model overfitting while obtaining well resolved peaks. Among the components picked, only the ones with a robust match with the NIST08 mass spectral library were retained in the final peak table, removing those coming from the baseline or column bleed. The models optimised by PARADISe using the 300 samples ended in a total of 107 components tentatively identified, recorded with their area in an *.xls* data table. Its capability for spectra deconvolution (even with unit mass resolution), for distinction between the signals from baseline or column bleeding, and for co-eluting components detection, allow PARADISe to get a consistent data matrix that simplifies the following steps, including the statistical analysis. For further information about PARADISe processing see (Sales et al., 2019).

3.3. Discriminant analysis by multivariate statistics

After internal standard normalisation and pareto-scaling of the peak data, the next step was the multivariate statistical analysis. After the aforementioned data transformation, the dataset was divided into two groups: 80 % of samples (considering each smoking treatment) for model training and the remaining 20 % for model validation. As an unbiased data exploratory analysis, a principal components analysis (PCA) was carried out considering the training data set. **Fig. 1** shows the score plot of the first two components of the same PCA, where PC1 and PC2 explain the 43.95 % and 9.94 % of the variance respectively, labelled by a) the two species of fish analysed and b) the different smoking treatments. From **Fig. 1a**, there was not significant inherent separation between the samples of the two fish species studied (tuna or swordfish). Thus, all the samples from the same treatment can be grouped in order to study the effect of the different treatments, regardless of the fish species. Regarding **Fig. 1b**, intrinsic

separation between the smoking treatments can be observed. The full cure (FCS) and medium (MCS) cold smoke treatment groups (CS) could be distinguished easily (green squares and light blue triangles, respectively) from the other samples (the non-cold smoke group, No-CS) along the first component (PC1). However, Light Cold Smoke (LCS) (dark blue triangles) could not be differentiated from the rest of the low to no treatment samples: Tasteless (TS), CO and Untreated fish (NAT).

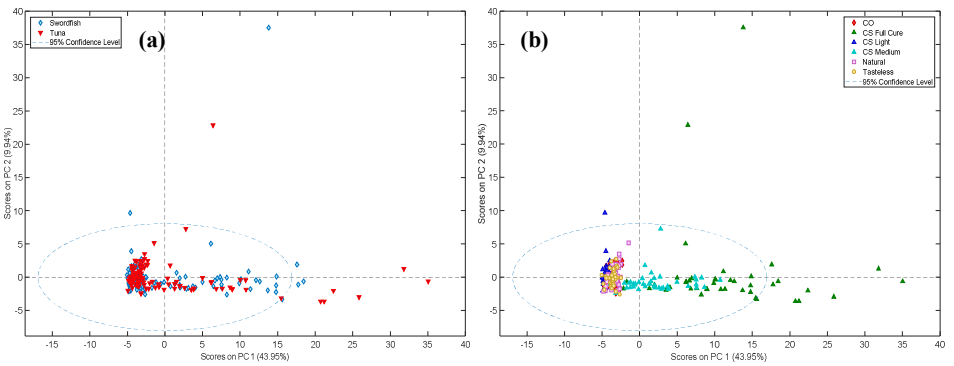


Fig. 1. *PCA score plots of the acquired data for the method training (a) coloured by fish species and (b) by smoking treatments.*

After PCA analysis, the partial least squares discriminant analysis (PLS-DA) was applied, which considers additional information about the groups to be classified (including the smoking treatment in the input information) (**Fig. 2**). The PLS-DA score plot of the latent variables 1 and 2 (LV1 vs LV2) (**Fig. 2a**) showed a clear separation between FCS and MCS from the remaining groups (and between them) along the LV1 axis. However, LCS group was still close to the non-cold smoke groups (TS, CO and NAT) although its differentiation improved regarding PCA since a gradual separation along LV2 could start to be observed. This distinction between the LCS and the No-CS groups was better noticed in the PLS-DA 3D (**Fig. 2b** and **c**).

In order to get information on relevant compounds related to the cold smoking process, PLS-DA between two groups was applied: the target class was “CS” (FCS, MCS and LCS) and the non-target class was “No-CS” (TS, CO and NAT). In order to verify the accuracy of the model, the classification of

the samples for the validation set was performed directly by the software. The PLS-DA was built using 4 latent variables (LV) and it explained 57.52 % variance. The classification plot obtained from this PLS-DA model is shown in **Fig. 3**, where samples were labelled by CS vs No-CS (**Fig. 3a**) and by the different smoking processes (**Fig. 3b**). It can be observed that all No-CS samples were at the same level, and they could be attributed a similar (and low) organoleptic load. Regarding the CS group, there is a slight variation depending on the time of exposure to smoke and therefore on the intensity level of the smoke flavour and odour, being the FCS samples at the top, MCS at lower level and LCS just above the threshold. This implies that LCS has low intensity organoleptic properties associated to cold smoke, but analysing the volatile composition it can be differentiated from other treatments that are not allowed in the European Union (TS and CO).

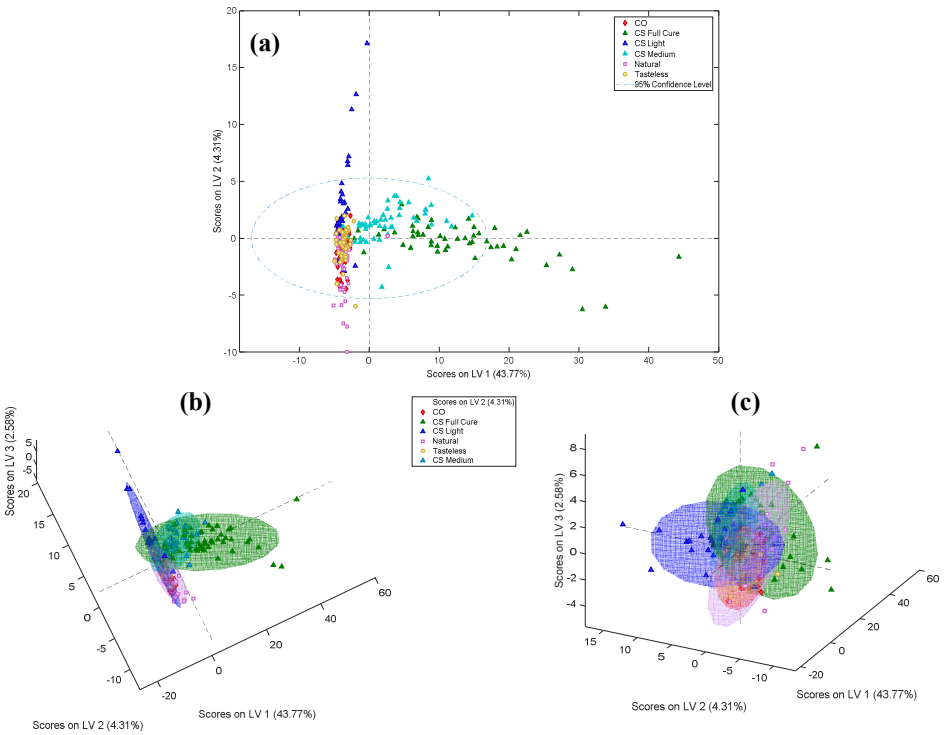


Fig. 2. *PLS-DA score plots of the acquired data for the method training (a) 2D plot plane LV1vs LV2 and (b) and (c) 3D score plots for LV1, LV2 and LV3.*

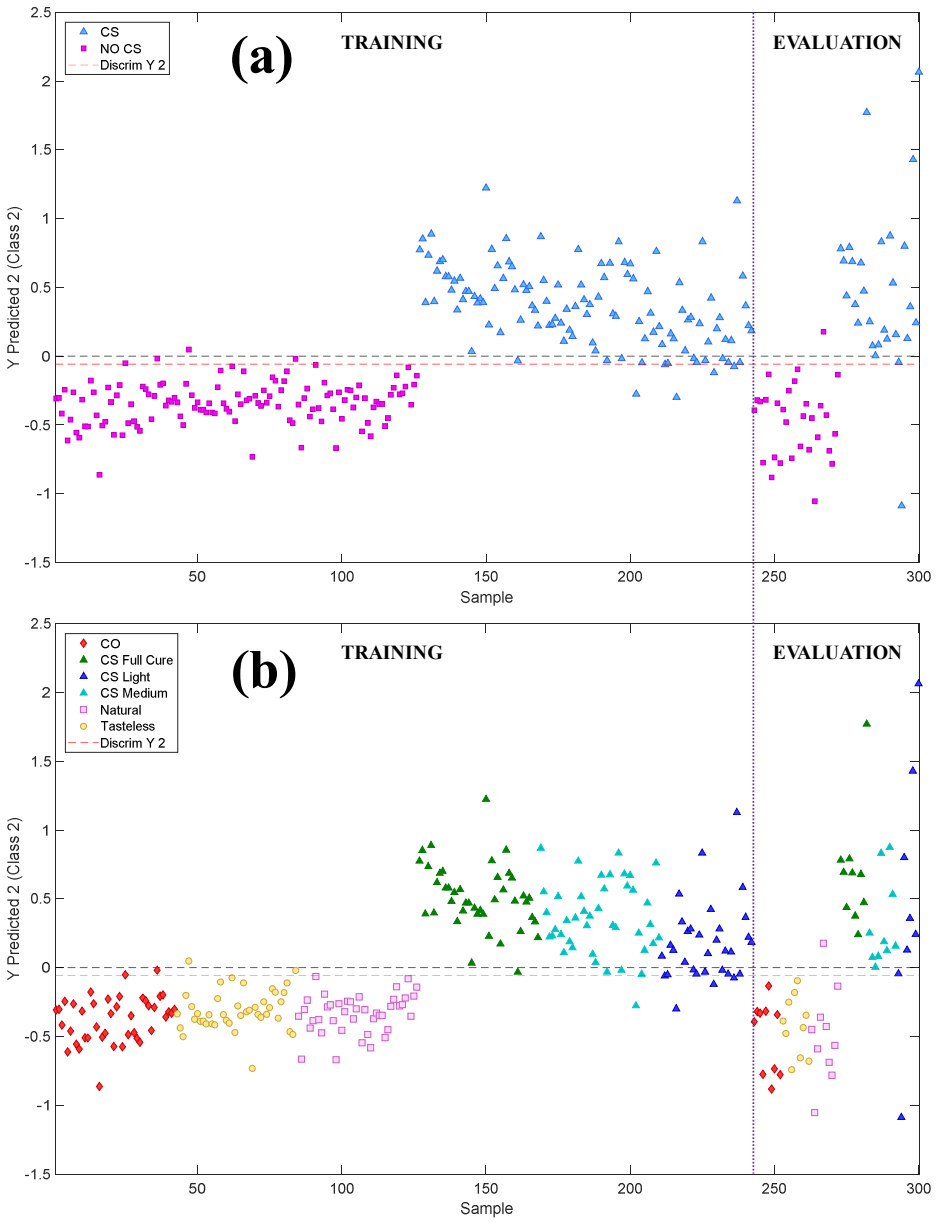


Fig. 3. *Model sample classification with the training and the evaluation set of samples coloured (a) by Cold smoked or non-cold smoked and (b) by the different smoking treatments.*

Confusion matrix summarizes the PLS-DA results of the training and validation set of samples (**Fig. 4)**. The model was able to correctly classify 100 % of the Full Cure cold smoke, the 97.6 % of the Medium cold smoke and 87.5 % of the Light cold smoke of the training samples. This model was then evaluated with the remaining 20 % of the samples, where 100 % of the Full Cure and Medium and 87.5 % of the Light cold smoked samples selected for evaluation were correctly classified as "Cold Smoked". It is worth noting that this model was highly efficient because all different CS treatments could be differentiated from other NO-CS processes that are forbidden in EU.

		OBJECTIVE	
		NO Cold Smoked	Cold Smoked
NO CS	CO	100 %	0
	Tasteless	100 %	0
	Natural	100 %	0
CS	CS Full Cure	0	100 %
	CS Medium	0	100 %
	CS Light	0	100 %

TRAINING SET				
	NO Cold Smoked		Cold Smoked	
CO	40/42	95.2 %	2/42	4.8 %
Tasteless	40/42	95.2 %	2/42	4.8 %
Natural	41/42	97.6 %	1/42	2.4 %
CS Full Cure	0/42	0 %	42/42	100 %
CS Medium	1/42	2.4 %	40/42	97.6 %
CS Light	4/32	12.5 %	28/32	87.5 %

EVALUATION SET				
	NO Cold Smoked		Cold Smoked	
CO	10/10	100 %	0/10	0 %
Tasteless	10/10	100 %	0/10	0 %
Natural	9/10	90 %	1/10	10 %
CS Full Cure	0/10	0 %	10/10	100 %
CS Medium	0/10	0 %	10/10	100 %
CS Light	1/8	0 %	7/8	87.5 %

Fig. 4. *Confusion matrix showing the comparison between the objective (up) of sample classification and the results (down) Training and evaluation set of samples after processing the samples through the entire developed procedure. The classification model was developed based on all the variables obtained (model score –0.05).*

3.4. Volatile profile-based classification model building and evaluation for different seafood treatments

The classification model was made with all 107 variables present in the samples. To generate a simpler model easier to apply in future targeted routine approach, most significant variables were selected according to the VIP score (Variable Importance in the Projection). This score summarizes the contribution a component makes to the PLS-DA model. In order to select as few variables as possible, the 29 features with VIP ≥ 1 were firstly selected to build the PLS-DA model. On the basis of the satisfactory results obtained, the

reduction of variables, applying the threshold VIP ≥ 1 for the new scores, continued progressively, until the model failed. Finally, the number of variables could be reduced down to 11 with the model still remaining consistent. The **Table 1** lists the main components selected as tentative markers for subsequent confirmation.

Table 1. *GC/MS measurements for the identified markers for the Cold smoke reduced classification model.*

VIP value	Marker	Molecular formula	M+• (*m/z*)	t_R (min)	LRI experimental (*)	LRI reported in NIST library (*)
2.50	3-methyl-cyclopentanone	$C_6H_{10}O$	98	10.86	1210	1218
2.33	Ethylbenzene	C_8H_{10}	106	8.51	1125	1122
1.94	2-methyl-2-cyclopenten-1-one	C_6H_8O	96	15.13	1366	1367
1.94	2-methyl-benzofuran	C_9H_8O	132	20.91	1593	1589
1.92	Furfuryl alcohol	$C_5H_6O_2$	98	22.46	1659	1659
1.85	2-acetylfuran	$C_6H_6O_2$	110	18.73	1504	1501
1.85	Acetophenone	C_8H_8O	120	22.24	1649	1649
1.67	Guaiacol	$C_7H_8O_2$	124	26.91	1863	1862
1.56	1-hydroxy-2-butanone	$C_4H_8O_2$	88	15.42	1376	1375
1.48	4-vinylguaiacol	$C_9H_{10}O_2$	150	30.06	(**)	(**)
1.01	Acetoin	$C_4H_8O_2$	88	12.93	1285	1285

(*) The Linear Retention Index (LRI) were obtained for each compound from NIST Library (https://webbook.nist.gov/) according to the most similar column and chromatographic conditions.

(**)The retention time fell out of the alkane range, the Linear Retention Index (LRI) could not be calculated.

Considering only the 11 tentative markers selected a new model was built and the classification plot is shown on **Fig. 5** and the confusion matrix on **Fig. 6**. With the 11 variables model, the group of samples are slightly (positively or negatively) affected by the variable reduction, being the Light cold smoke the most affected (81.3 % were correctly classified in the training samples in comparison to the 87.5 % of the previous model). Nevertheless, this is still an acceptable result and it allowed a satisfactory classification of the samples randomly selected for evaluation. Thus, 100 % FCS and MCS, and 87.5 % of the LCS samples, were satisfactorily classified as "Cold Smoke" samples with the developed model, reaching the same results as for the 107-classification model.

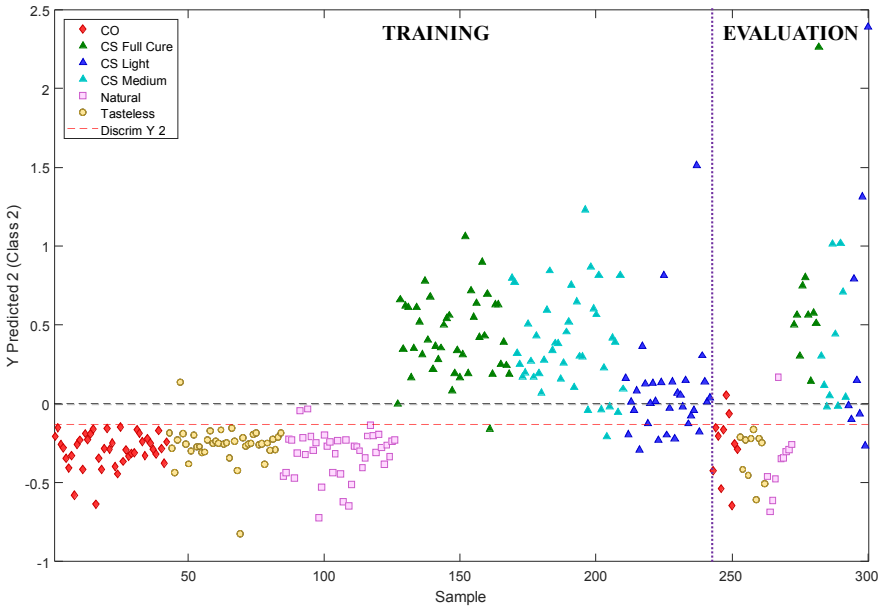


Fig. 5. *Model sample classification with 11 variables with the training and the evaluation set of samples coloured by the different smoking treatments.*

		OBJECTIVE	
		NO Cold Smoked	Cold Smoked
NO CS	CO	100 %	0
	Tasteless	100 %	0
	Natural	100 %	0
CS	CS Full Cure	0	100 %
	CS Medium	0	100 %
	CS Light	0	100 %

TRAINING SET				
	NO Cold Smoked		Cold Smoked	
CO	42/42	100 %	0/42	0 %
Tasteless	41/42	97.6 %	1/42	2.4 %
Natural	40/42	95.2 %	2/42	4.8 %
CS Full Cure	1/42	2.4 %	41/42	97.6 %
CS Medium	1/42	2.4 %	40/42	97.6 %
CS Light	6/32	18.7 %	26/32	81.3 %

EVALUATION SET				
	NO Cold Smoked		Cold Smoked	
CO	8/10	80 %	2/10	20 %
Tasteless	10/10	100 %	0/10	0 %
Natural	9/10	90 %	1/10	10 %
CS Full Cure	0/10	0 %	10/10	100 %
CS Medium	0/10	0 %	10/10	100 %
CS Light	1/8	12.5 %	7/8	87.5 %

Fig. 6. *Confusion matrix showing the comparison between the objective (up) of sample classification and the results (down) Training and evaluation set of samples after processing the samples through the entire developed procedure. The classification model was developed based on the 11 most relevant variables obtained (model score –0.1).*

3.5. Cold smoke-related compounds identification in the simplified classification model

Identification of the 11 compounds used in the simplified model is crucial for the development of targeted methods in future works, and for validation of the classification model by analysing a large amount of samples and blind samples prior to its routine analysis application.

An interesting contribution of PARADISe software is that automatically performs a comparison between the deconvoluted spectra and the NIST EI mass spectra library (in this case NIST08) giving the best fitted candidate. In order to increase the confidence in the identification, Linear Retention Indices (LRIs) were calculated for each compound using a C7-C20 alkane mixture. The tentative identification for a compound was assigned when the NIST match for this compound was over 800; and the LRI match with the NIST library was below ± 20. Finally, the identity of the 11 markers was confirmed by the injection of reference standards under sample identical conditions, and subsequent comparison of spectra and retention times, according to the criteria of our laboratory and the *Chemical Analysis Working group (CAWG) Metabolomics Standards Initiative (MSI)* (Sumner et al., 2007). Results are shown in **Table 1** with their molecular formula, the detected molecular ion, the NIST match, retention time and LRI. In the case of the 4-vinylguaiacol, the retention time fell out of the C20 alkane range, thus the LRI could not be calculated. However, its identity was confirmed with the reference standard injection. Our results were compared with the available literature in the field of organoleptic cold smoke composition, which reports that ketones, phenolic derivatives and furan derivative compounds are associated with smoke flavour and odour (Gómez-Estaca, Gómez-Guillén, Montero, Sopelana, & Guillén, 2011; M.D. Guillén et al., 2006, Jónsdóttir et al., 2008). The compound with higher importance in our 11 variables classification model was 3-methyl-cyclopentanone, previously identified as wood smoke component together with other ketones such as 2-methyl-2-cyclopenten-1-one, 3-hydroxy-2-butanone, 1-hydroxy-2-butanone and acetophenone (Vidal et al., 2017). Guaiacol and guaiacol derivatives, such as 4-vinyl guaiacol has been previously reported as derived from wood pyrolysis and with a high importance in the smoke flavour grade. Associated with this

property are also furan derivatives compounds, such as 2-acetylfuran, 2-methyl-benzofuran and 2-furanmethanol (Jónsdóttir et al., 2008). Finally, hydrocarbons, like ethylbenzene, part of the wood smoke constituents (María D. Guillén & Errecalde, 2002) were also found as marker although no references about its flavour contribution was found in the literature.

4. Conclusions

The use of untargeted volatolomics based on DHS-P&T for volatile extraction and subsequent analysis by GC/MS has allowed to obtain relevant information regarding the volatile composition of smoked fish samples, responsible for the classification of smoking technique applied. The use of PARADISe has allowed robust peak detection, cleaner spectra and, in combination with NIST libraries, an efficient tentative identification. Using this methodology a classification model has been developed able to distinguish samples with "Cold Smoked" treatment (Full Cure, Medium and Light smoked) from those without "Cold Smoked" treatment (Tasteless, CO and untreated), and it has allowed to build a consistent statistical model for correct classification. The model built with all the 107 detected compounds allowed the correct classification of 96.3 % of the blind samples, while using the simplified model, based on only 11 identified compounds, 95 % of the blind samples were still correctly classified. The confirmation of the identity of these 11 compounds with their reference standards will allow in the near future to develop a targeted method to be implemented in routine analysis.

The model was developed to classify fish flesh treated with cold smoking at 4 °C and other non-cold smoked treatments. In this work, relative chromatographic areas, referred to internal standard, were used for classification purposes. For future routine applications, reference standards should be used in order to calculate the concentration of each marker in the samples, following a simple target quantitative approach.

The possibility to apply the developed methodology to other fish flesh samples treated with cold smoking is promising but should be further studied and validated. Thus, the markers proposed for the cold smoked seafood analysed in this work may not be the most suitable for other cold smoking processes performed by other companies. This is due to the variability

between the cold smoke treatments applied (type of wood, exposure time, temperature, salinity…).

CRediT authorship contribution statement

Leticia Lacalle-Bergeron: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Funding acquisition. **Tania Portolés**: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Carlos Sales**: Methodology, Software, Formal analysis, Investigation, Visualization. **M. Carmen Corell**: Conceptualization, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration, Funding acquisition. **Fernando Domínguez**: Conceptualization, Resources, Writing - review & editing, Visualization, Project administration. **Joaquim Beltrán**: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft. **Juan Vicente Sancho:** Conceptualization, Methodology, Formal analysis, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Félix Hernández**: Validation, Resources, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

L Lacalle-Bergeron acknowledges the financial support of Universitat Jaume I, Spain for his pre-doctoral grant (UJI19I001/03). T. Portoles acknowledges Ramon y Cajal Program from the Ministry of Economy and Competitiveness, Spain (RYC-2017-22525) for funding her research. The authors acknowledge the financial support of Generalitat Valenciana, as research group of excellence PROMETEO/2019/040.

References

Abou-el-karam et al., 2017, S. Abou-el-karam, J. Ratel, N. Kondjoyan, C. Truan, E. Engel, Marker discovery in volatolomics based on systematic alignment of GC-MS signals: Application to food authentication, Analytica Chimica Acta, 991 (2017), pp. 58-67, 10.1016/j.aca.2017.08.019

Barstad et al., 2006, H. Barstad, A.C. Alvik, E. Løvaas, Seafood research from fish to dish, J.B. Luten, C. Jacobsen, K. Bekaert, A. Sæbø, J. Oehlenschläger (Eds.), Seafood research from fish to dish: Quality, safety and processing of wild and farmed seafood, Wageningen Academic Publishers (2006), 10.3920/978-90-8686-581-9

Bartolucci et al., 2010, G. Bartolucci, E. Droghetti, C. Focardi, M. Bambagiotti-Alberti, M. Nocentini, G. Smulevich, High throughput headspace GC-MS quantitative method to measure the amount of carbon monoxide in treated tuna fish, Journal of Mass Spectrometry, 45 (9) (2010), pp. 1041-1045, 10.1002/jms.1798

Beaulieu and Lea, 2006, J.C. Beaulieu, J.M. Lea, Characterization and semiquantitative analysis of volatiles in seedless watermelon varieties using solid-phase microextraction, Journal of Agricultural and Food Chemistry, 54 (20) (2006), pp. 7789-7793, 10.1021/jf060663l

Beltran et al., 2006, J. Beltran, E. Serrano, F.J. López, A. Peruga, M. Valcarcel, S. Rosello, Comparison of two quantitative GC–MS methods for analysis of tomato aroma based on purge-and-trap and on solid-phase microextraction, Analytical and Bioanalytical Chemistry, 385 (7) (2006), pp. 1255-1264, 10.1007/s00216-006-0410-9

Ben Brahim et al., 2018, S. Ben Brahim, A. Amanpour, F. Chtourou, H. Kelebek, S. Selli, M. Bouaziz, Gas chromatography-mass spectrometry–olfactometry to control the aroma fingerprint of extra virgin olive oil from three tunisian cultivars at three harvest times, Journal of Agricultural and Food Chemistry, 66 (11) (2018), pp. 2851-2861, 10.1021/acs.jafc.7b06090

Bouhlel et al., 2017, J. Bouhlel, J. Ratel, S. Abouelkaram, F. Mercier, A. Travel, E. Baéza, Catherine Jondreville, Gaud Dervilly-Pinel, Philippe Marchand, Bruno Le Bizec, Estelle Dubreil, Sophie Mompelat, Eric Verdon, Chanthadary Inthavong, Thierry Guérin, Douglas N. Rutledge, E. Engel, Solid-phase microextraction set-up for the analysis of liver volatolome to detect livestock exposure to micropollutants, Journal of Chromatography A, 1497 (2017), pp. 9-18, 10.1016/j.chroma.2017.03.008

Broza et al., 2015, Y.Y. Broza, P. Mochalski, V. Ruzsanyi, A. Amann, H. Haick, Hybrid volatolomics and disease detection, Angewandte Chemie International Edition, 54 (38) (2015), pp. 11036-11048, 10.1002/anie.201500153

Dalgaard et al., 2008, Dalgaard, P., Emborg, J., Kjølby, A., Sørensen, N. D., & Ballin, N. Z. (2008). Histamine and biogenic amines: formation and importance in seafood. In Improving seafood products for the consumer (pp. 292–324). Elsevier. 10.1533/9781845694586.3.292

Dettmer et al., 2007, K. Dettmer, P.A. Aronov, B.D. Hammock, Mass spectrometry-based metabolomics, Mass Spectrometry Reviews, 26 (1) (2007), pp. 51-78, 10.1002/mas.20108

Dirinck et al., 1977, P. Dirinck, L. Schreyen, N. Schamp, Aroma quality evaluation of tomatoes, apples, and strawberries, Journal of Agricultural and Food Chemistry, 25 (4) (1977), pp. 759-763, 10.1021/jf60212a039

Dudzik et al., 2018, D. Dudzik, C. Barbas-Bernardos, A. García, C. Barbas, Quality assurance procedures for mass spectrometry untargeted metabolomics, a review, Journal of Pharmaceutical and Biomedical Analysis, 147 (2018), pp. 149-173, 10.1016/j.jpba.2017.07.044

Dudzik et al., 2017, D. Dudzik, M. Zorawski, M. Skotnicki, W. Zarzycki, A. García, S. Angulo, M. Paz Lorenzo, Coral Barbas, M.P. Ramos, GC–MS based gestational diabetes mellitus longitudinal study: Identification of 2- and 3-hydroxybutyrate as potential prognostic biomarkers, Journal of Pharmaceutical and Biomedical Analysis, 144 (2017), pp. 90-98, 10.1016/j.jpba.2017.02.056

Fredes et al., 2016, A. Fredes, C. Sales, M. Barreda, M. Valcárcel, S. Roselló, J. Beltrán, Quantification of prominent volatile compounds responsible for muskmelon and watermelon aroma by purge and trap extraction followed by gas chromatography–mass spectrometry determination, Food Chemistry, 190 (2016), pp. 689-700, 10.1016/j.foodchem.2015.06.011

Garcia and Barbas, 2011, Garcia, A., & Barbas, C. (2011). Gas chromatography-mass spectrometry(GC-MS)-based metabolomics. In T. O. Metz (Ed.), Metabolic profiling, methods and protocols (Vol. 708, pp. 191–204), 10.1007/978-1-61737-985-7

Gil-Solsona et al., 2016, R. Gil-Solsona, M. Raro, C. Sales, L. Lacalle, R. Díaz, M. Ibáñez, Joaquim Beltran, Juan Vicente Sancho, F.J. Hernández, Metabolomic approach for Extra virgin olive oil origin discrimination making use of ultra-high performance liquid chromatography – Quadrupole time-of-flight mass spectrometry, Food Control, 70 (2016), pp. 350-359, 10.1016/j.foodcont.2016.06.008

Gokoglu, 2020, N. Gokoglu, Innovations in seafood packaging technologies: A review, Food Reviews International, 36 (4) (2020), pp. 340-366, 10.1080/87559129.2019.1649689

Gómez-Estaca et al., 2011, J. Gómez-Estaca, M.C. Gómez-Guillén, P. Montero, P. Sopelana, M.D. Guillén, Oxidative stability, volatile components and polycyclic aromatic hydrocarbons of cold-smoked sardine (*Sardina pilchardus*) and dolphinfish (*Coryphaena hippurus*), LWT - Food Science and Technology, 44 (6) (2011), pp. 1517-1524, 10.1016/j.lwt.2011.02.006

Guillén et al., 2006, M.D. Guillén, M.C. Errecalde, J. Salmerón, C. Casas, Headspace volatile components of smoked swordfish (*Xiphias gladius*) and cod (*Gadus morhua*) detected by means of solid phase microextraction and gas chromatography–mass spectrometry, Food Chemistry, 94 (1) (2006), pp. 151-156, 10.1016/j.foodchem.2005.01.014

Guillén and Errecalde, 2002, María D. Guillén, M.C. Errecalde, Volatile components of raw and smoked black bream (*Brama raii*) and rainbow trout (*Oncorhynchus mykiss*) studied by means of solid phase

microextraction and gas chromatography/mass spectrometry, Journal of the Science of Food and Agriculture, 82 (9) (2002), pp. 945-952, 10.1002/jsfa.1128

Harshman, 1972, R.A. Harshman, PARAFAC2: Mathematical and technical notes, UCLA Working Papers in Phonetics, 22 (10) (1972), pp. 30-44

Huang et al., 2019, X.H. Huang, L.B. Qi, B.S. Fu, Z.H. Chen, Y.Y. Zhang, M. Du, Xiu-Ping Dong, Bei-Wei Zhu, L. Qin, Flavor formation in different production steps during the processing of cold-smoked Spanish mackerel, Food Chemistry, 286 (2019), pp. 241-249, 10.1016/j.foodchem.2019.01.211

Johnsen et al., 2014, L.G. Johnsen, J.M. Amigo, T. Skov, R. Bro, Automated resolution of overlapping peaks in chromatographic data, Journal of Chemometrics, 28 (2) (2014), pp. 71-82, 10.1002/cem.2575

Johnsen et al., 2017, L.G. Johnsen, P.B. Skou, B. Khakimov, R. Bro, Gas chromatography – mass spectrometry data processing made easy, Journal of Chromatography A, 1503 (2017), pp. 57-64, 10.1016/j.chroma.2017.04.052

Jónsdóttir et al., 2008, R. Jónsdóttir, G. Ólafsdóttir, E. Chanie, J.E. Haugen, Volatile compounds suitable for rapid detection as quality indicators of cold smoked salmon (*Salmo salar*), Food Chemistry, 109 (1) (2008), pp. 184-195, 10.1016/j.foodchem.2007.12.006

Khakimov et al., 2016, B. Khakimov, R.J. Mongi, K.M. Sørensen, B.K. Ndabikunze, B.E. Chove, S.B. Engelsen, A comprehensive and comparative GC–MS metabolomics study of non-volatiles in Tanzanian grown mango, pineapple, jackfruit, baobab and tamarind fruits, Food Chemistry, 213 (2016), pp. 691-699, 10.1016/j.foodchem.2016.07.005

Marušić Radovčić et al., 2016, N. Marušić Radovčić, S. Vidaček, T. Janči, H. Medić, Characterization of volatile compounds, physico-chemical and sensory characteristics of smoked dry-cured ham, Journal of Food Science and Technology, 53 (11) (2016), pp. 4093-4105, 10.1007/s13197-016-2418-2

Płotka-Wasylka et al., 2017, J. Płotka-Wasylka, N. Szczepańska, K. Owczarek, J. Namieśnik, Miniaturized solid phase extraction, Comprehensive Analytical Chemistry, 76 (2017), pp. 279-318, 10.1016/bs.coac.2017.03.001

Saldaña et al., 2019, Saldaña, E., Saldarriaga, L., Cabrera, J., Siche, R., Behrens, J. H., Selani, M. M., de Almeida, Luciana Duque Silva, Jair S. Silva Pinto, Contreras-Castillo, C. J. (2019). Relationship between volatile compounds and consumer-based sensory characteristics of bacon smoked with different Brazilian woods. Food Research International, 119(September 2018), 839–849, 10.1016/j.foodres.2018.10.067

Sales et al., 2017, C. Sales, M.I. Cervera, R. Gil, T. Portolés, E. Pitarch, J. Beltran, Quality classification of Spanish olive oils by untargeted gas chromatography coupled to hybrid quadrupole-time of flight mass spectrometry with atmospheric pressure chemical ionization and metabolomics-based statistical approach, Food Chemistry, 216 (2017), pp. 365-373, 10.1016/j.foodchem.2016.08.033

Sales et al., 2019, C. Sales, T. Portolés, L.G. Johnsen, M. Danielsen, J. Beltran, Olive oil quality classification and measurement of its organoleptic attributes by untargeted GC–MS and multivariate statistical-based approach, Food Chemistry, 271 (2019), pp. 488-496, 10.1016/j.foodchem.2018.07.200

Sérot et al., 2004, T. Sérot, R. Baron, C. Knockaert, J.L. Vallet, Effect of smoking processes on the contents of 10 major phenolic compounds in smoked fillets of herring (*Cuplea harengus*), Food Chemistry, 85 (1) (2004), pp. 111-120, 10.1016/j.foodchem.2003.06.011

Soria et al., 2017, Ana C. Soria, M.J. García-Sarrió, A.I. Ruiz-Matute, M.L. Sanz, Headspace techniques for volatile sampling, Comprehensive Analytical Chemistry, 76 (2017), pp. 255-278, 10.1016/bs.coac.2017.02.001

Soria et al., 2008, Ana Cristina Soria, I. Martínez-Castro, J. Sanz, Some aspects of dynamic headspace analysis of volatile components in honey, Food Research International, 41 (8) (2008), pp. 838-848, 10.1016/j.foodres.2008.07.010

Sumner et al., 2007, L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, Teresa W.-M. Fan, Oliver Fiehn, Royston Goodacre, Julian L. Griffin, Thomas Hankemeier, Nigel Hardy, James Harnly, Richard Higashi, Joachim Kopka, Andrew N. Lane, John C. Lindon, Philip Marriott, Andrew W. Nicholls, Michael D. Reily, John J. Thaden, M.R. Viant, Proposed minimum reporting standards for chemical analysis, Metabolomics, 3 (3) (2007), pp. 211-221, 10.1007/s11306-007-0082-2

Thomsen et al., 2016, B.R. Thomsen, B. Yesiltas, A.D.M. Sørensen, D.B. Hermund, J. Glastrup, C. Jacobsen, Comparison of three methods for extraction of volatile lipid oxidation products from food matrices for GC–MS analysis, JAOCS, Journal of the American Oil Chemists' Society, 93 (7) (2016), pp. 929-942, 10.1007/s11746-016-2837-2

Tomita et al., 2018, S. Tomita, T. Nakamura, S. Okada, NMR- and GC/MS-based metabolomic characterization of *sunki*, an unsalted fermented pickle of turnip leaves, Food Chemistry, 258 (2018), pp. 25-34, 10.1016/j.foodchem.2018.03.038

van den Berg et al., 2006, R.A. van den Berg, H.C.J. Hoefsloot, J.A. Westerhuis, A.K. Smilde, M.J. van der Werf, Centering, scaling, and transformations: Improving the biological information content of metabolomics data, BMC Genomics, 7 (2006), 10.1186/1471-2164-7-142

Varlet et al., 2009, V. Varlet, T. Serot, C. Prost, Smoke flavoring technology in seafood, Handbook of seafood and seafood products analysis, CRC Press (2009), pp. 233-254, 10.1201/9781420046359-c15

Vidal et al., 2017, N.P. Vidal, E. Goicoechea, M.J. Manzanos, M.D. Guillén, Effect of smoking using smoke flavorings on several characteristics of farmed sea bass (*Dicentrarchus labrax*) fillets and on their evolution during vacuum-packed storage at refrigeration temperature, Journal of Food Processing and Preservation, 41 (2) (2017), Article e12800, 10.1111/jfpp.12800

Supplementary Information

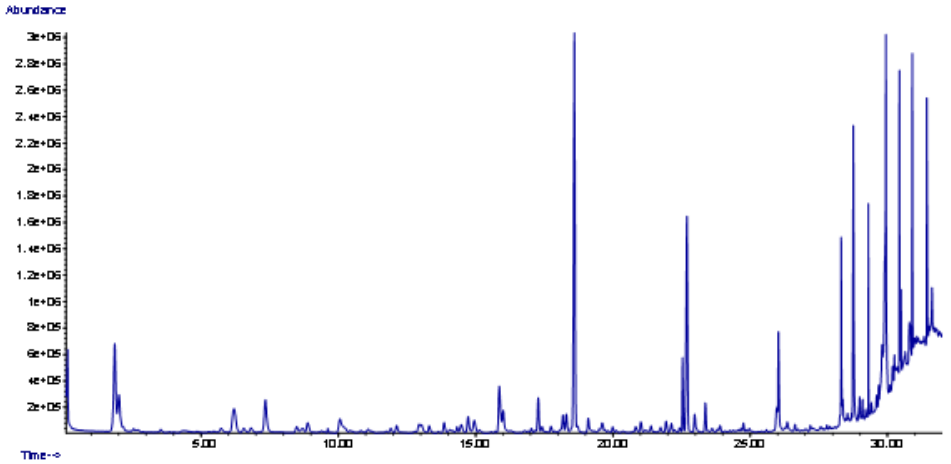


Fig. S2. *Total ion chromatogram from the volatile extract of a light cure cold smoked tuna sample obtained by Dynamic Head Space-Purge and Trap (DHS-P&T), followed by the injection with a thermal desorption unit (TDU) and CIS 4 PTV injector, under the chromatographic conditions described in section 2.4.*

2.3 Discussion of the results

Considering the different groups of compounds that could be addressed by untargeted metabolomics and the expected characteristics from smoked products, **Scientific Article II** focused on the most appreciated aspect by consumers, the aroma. This was expected to be the biggest differentiation between cold smoked products and carbon monoxide (CO) or filtered smoke treated products, since both latter treatments try to maintain the fish natural flavour. The chemicals related to the sensory properties of smoked products are volatile compounds acquired from the wood smoke, some of them with reported microbistatic activity. CO and filtered smoke treatments lack this added property that has made the smoking process one of the oldest food preservation and safety techniques.

However, the consumer's search for new experiences and flavours has led to the production of different smoked products with different flavour intensities, some of them so mild that they are difficult to distinguish from the two aforementioned treatments banned by European Union, and even untreated raw fish, by traditional methodologies employed in trade controls.

For this reason, DHS-P&T was selected as extraction method, since it is characterised by its high preconcentration factor of volatiles present in the sample and it allows the subsequent detection of compounds even in lightly flavoured smoked fishery products.

This work was not only focused on the determination of the characteristic compounds of the cold smoking process, but a classification model was also developed based on 11 compounds with confirmed identity that allowed the correct classification of 95 % of the blind samples.

For the selection of the compounds to be used in the model building, the features significantly different highlighted in the statistics were evaluated one by one. One of the problems found in the selection was related to the peak picking process carried out by the PARADISe software, although it is not characteristic of this software in particular, but a general problem of peak picking process in metabolomics. The peak picking was done with samples from all the groups to be compared at the same time, in order to correlate the

peaks from one sample to another. Some samples were characterised by a series of volatiles at very different concentrations, the full cure cold smoked group being very different in terms of concentration of certain compounds even with the following group in flavour intensity, medium cold smoke. The problem arises when the program tries to match those peaks with high intensity in one group to another one where they are hardly detectable or even non-existent. In these cases, the program can deconvolute a peak when it is actually not present. Moreover, there is possibly an overfitting effect in discriminant statistical analysis if a group is constituted from very disparate samples in terms of concentration. This combination may result in the highlighting of some compounds that are not really present in all the samples of the conformed group. These facts added some difficulty in the selection of compounds that were specific to the cold smoking treatment and not from the different smoking degrees.

Currently, it is being studied the transfer of the classification model to an analysis method easily applicable to routine laboratories. DHS-P&T is a very efficient technique for scanning the compounds present in the sample for a global screening approach as untargeted metabolomics, but it has been shown to be an extraction technique difficult to optimise for quantitative methods. Moreover, the extraction is performed off-line from the GC-MS analysis and there is no automation in the extraction process. These facts make necessary an adaption of the method.

In this ongoing targeted method, the analysis of the 11 highlighted compounds in the untargeted metabolomics will be sought, as well as other compounds that the European Union have included in the regulation regarding the smoked food products. Contrary to the more universal sample treatment trend from untargeted metabolomics, parameters around the extraction are being optimised for the enhancement of the compounds selected in the method. For this purpose and counting on the automation of different types of on-line volatiles extraction offered by our GC-EI-HRMS, HS-SPME with different fibres as well as the in-tube extraction dynamic headspace (ITEX-DHS) are being explored.

Although this methodology has being developed in collaboration with Sea Delight Europe, SL, it is also planned to validate the selected markers

with fishery products from other companies in which, within the cold smoking procedure, certain parameters could vary, giving the final product differentiating flavours.

2.4 References

Brigante, F. I., Podio, N. S., Wunderlin, D. A., & Baroni, M. V. (2022). Comparative metabolite fingerprinting of chia, flax and sesame seeds using LC-MS untargeted metabolomics. Food Chemistry, 371(April 2021), 131355. https://doi.org/10.1016/j.foodchem.2021.131355

Castro-Puyana, M., Pérez-Míguez, R., Montero, L., & Herrero, M. (2017). Application of mass spectrometry-based metabolomics approaches for food safety, quality and traceability. TrAC Trends in Analytical Chemistry, 93, 102–118. https://doi.org/10.1016/j.trac.2017.05.004

Giannenas, I., Grigoriadou, K., Sidiropoulou, E., Bonos, E., Cheilari, A., Vontzalidou, A., Karaiskou, C., Aligiannis, N., Florou-Paneri, P., & Christaki, E. (2021). Untargeted UHPLC-MS metabolic profiling as a valuable tool for the evaluation of eggs quality parameters after dietary supplementation with oregano, thyme, sideritis tea and chamomile on brown laying hens. Metabolomics, 17(6), 51. https://doi.org/10.1007/s11306-021-01801-7

Jacobs, D. M., van den Berg, M. A., & Hall, R. D. (2021). Towards superior plant-based foods using metabolomics. Current Opinion in Biotechnology, 70, 23–28. https://doi.org/10.1016/j.copbio.2020.08.010

Li, R., Sun, Z., Zhao, Y., Li, L., Yang, X., Cen, J., Chen, S., Li, C., & Wang, Y. (2021). Application of UHPLC-Q-TOF-MS/MS metabolomics approach to investigate the taste and nutrition changes in tilapia fillets treated with different thermal processing methods. Food Chemistry, 356(January), 129737. https://doi.org/10.1016/j.foodchem.2021.129737

Lozano-Castellón, J., Rocchetti, G., Vallverdú-Queralt, A., Illán, M., Torrado-Prat, X., Lamuela-Raventós, R. M., & Lucini, L. (2021). New vacuum cooking techniques with extra-virgin olive oil show a better phytochemical profile than traditional cooking methods: A foodomics study. Food Chemistry, 362(April), 130194. https://doi.org/10.1016/j.foodchem.2021.130194

Nyarukowa, C., van Reenen, M., Koech, R., Mason, S., Lindeque, Z., Kamunya, S., Mose, R., & Apostolides, Z. (2021). Untargeted metabolomics

reveals differences between commercial and non-commercial Camellia sinensis cultivars used in black tea production. Journal of Plant Biochemistry and Biotechnology, 9. https://doi.org/10.1007/s13562-021-00722-9

Qiu, S., Tu, Y., Huang, D., Dong, Z., Huang, M., Cheng, J., Tan, J., Chen, W., Sun, L., & Chen, W. (2021). Selection of appropriate post-harvest processing methods based on the metabolomics analysis of Salvia miltiorrhiza Bunge. Food Research International, 144(April), 110366. https://doi.org/10.1016/j.foodres.2021.110366

Rangel-Huerta, O. D., Uhlig, S., Ivanova, L., Dang, T. T., Rode, T. M., Noriega Fernández, E., & Fæste, C. K. (2022). Metabolomics workflow for quality control of differently-processed pre-cooked chicken fillets. Food Chemistry, 370(April 2021), 131006. https://doi.org/10.1016/j.foodchem.2021.131006

Rivera-Pérez, A., Romero-González, R., & Garrido Frenich, A. (2021). Application of an innovative metabolomics approach to discriminate geographical origin and processing of black pepper by untargeted UHPLC-Q-Orbitrap-HRMS analysis and mid-level data fusion. Food Research International, 150, 110722. https://doi.org/10.1016/j.foodres.2021.110722

Wang, K., & Xu, Z. (2022). Comparison of freshly squeezed, Non-thermally and thermally processed orange juice based on traditional quality characters, untargeted metabolomics, and volatile overview. Food Chemistry, 373(PB), 131430. https://doi.org/10.1016/j.foodchem.2021.131430

Chapter 3.
Food intake

Chapter 3. Food intake

3.1 Introduction

3.2 Scientific article III

"Ultra-performance liquid chromatography - ion mobility separation – quadrupole time-of-flight MS (UHPLC-IMS-QTOF MS) metabolomics for short-term biomarker discovery of orange intake: a randomized, controlled study"

Nutrients 2020,12,1916; doi:10.3390/nu12071916

3.3 Medium-term biomarkers of orange intake

3.4 References

3.1 Introduction

Diet is one of the environmental exposures with higher impact on health, and therefore, its study is vital for the health-related sciences. As explained previously in this thesis, the measurement of dietary exposure has been carried out traditionally by means of collecting tool information based on food frequency questionnaires (FFQ), food diaries or 24-h recalls. All of them are based on the information given by the patient/subject, often underreported, with recall error or unreliable data sometimes due to the social desirability bias. The measurement of intake and diet should be both accurate and capable of being applied to a large number of subjects, requirements that, more often than not, these traditional techniques are unable to meet (O'Gorman & Brennan, 2017).

The metabolome responds to stress, disease or nutrients long before the alteration of proteome or transcriptome, making the study of metabolic alteration very attractive for multiple fields, including dietary intake assessment. Thanks to the untargeted metabolomics, a response is being given in this area by the discovery of compounds specific for food or group of foods that become bioavailable after their consumption, biomarkers of food intake (BFI). Once the biomarkers are determined, they could be used for diet monitoring in a precise and objective way; as well as to contrast the information obtained by more traditional methods, and hence better unravel the associations between diet and health (Vázquez-Manjarrez et al., 2020).

Around this objective a series of initiatives are being created; as the Food Biomarker Alliance (FoodBAll, *www.foodmetabolome.org*) which is a multicentric project aiming to develop strategies for food biomarker discovery and validation, as well as specialised data bases as FooDB (*www.foodb.ca*), PhytoHub (*www.phytohub.eu*) or FoodComEx (www.foodcomex.org), with close collaboration with other analytical or specialised data bases (MassBank *www.massbank.eu* , Exposome Explorer *www.exposome-explorer.iarc.fr* …) (Vázquez-Manjarrez et al., 2020). Thanks to this type of initiatives, a great source of knowledge is being built in this field, that will allow the implementation of monitoring tools to obtain more reliable data.

Moreover, untargeted metabolomics in nutrition research has been also applied to dietary patterns (DPs) (Wu et al., 2021). There is a great diversity of DPs based on multiple factors such as the region or country (e.g. Mediterranean or Nordic diet), on focusing in a determinate type of food (e.g. vegetarian or vegan diet), designed to a specific subject (e.g. DASH diet o Portfolio diet to lower blood pressure and cholesterol levels respectively), to a determinate goal (low-carbohydrates diet) or even the trends of the moment (e.g. paleo diet, Dunkan diet…) (Gomez-Delgado et al., 2021). Their monitoring is therefore based on a multiple biomarker depending on the foods employed in the diet under study (Kim & Rebholz, 2021).

In this chapter, untargeted metabolomic approach has been applied in the discovery of short-term biomarkers of orange intake (**Scientific Article III**) and an exploratory study for medium-term biomarker (**Section 3.3**), a citrus fruit with great importance in the Mediterranean diet and widely consumed around the world by itself or derived food products as juices. In addition, it is an example of how the development of new technologies facilitates the application of untargeted metabolomic approaches, not only in this area but in all its fields of application, specifically the introduction of ion mobility separation (IMS) to the HRMS-based instruments.

3.2 Scientific article III

nutrients

MDPI

Article

Ultra-Performance Liquid Chromatography-Ion Mobility Separation-Quadruple Time-of-Flight MS (UHPLC-IMS-QTOF MS) Metabolomics for Short-Term Biomarker Discovery of Orange Intake: A Randomized, Controlled Crossover Study

Leticia Lacalle-Bergeron [1], Tania Portolés [1], Francisco J. López [1], Juan Vicente Sancho [1], Carolina Ortega-Azorín [2,3], Eva M. Asensio [2,3], Oscar Coltell [3,4] and Dolores Corella [2,3,*]

1 Research Institute for Pesticides and Water (IUPA), Universitat Jaume I, 12071 Castellón, Spain; mlacalle@uji.es (L.L.-B.); tania.portoles@qfa.uji.es (T.P.); lopezf@uji.es (F.J.L.); sanchoj@uji.es (J.V.S.)
2 Department of Preventive Medicine and Public Health, School of Medicine, University of Valencia, 46010 Valencia, Spain; Carolina.Ortega@uv.es (C.O.-A.); eva.m.asensio@uv.es (E.M.A.)
3 CIBER Fisiopatología de la Obesidad y Nutrición (CIBEROBN), Instituto de Salud Carlos III, 28029 Madrid, Spain; oscar.coltell@uji.es
4 Department of Computer Languages and Systems, Universitat Jaume I, 12071 Castellón, Spain
* Correspondence: dolores.corella@uv.es; Tel.: +34-963-86-4800

Received: 26 May 2020; Accepted: 25 June 2020; Published: 29 June 2020

check for updates

Abstract: A major problem with dietary assessments is their subjective nature. Untargeted metabolomics and new technologies can shed light on this issue and provide a more complete picture of dietary intake by measuring the profile of metabolites in biological samples. Oranges are one of the most consumed fruits in the world, and therefore one of the most studied for their properties. The aim of this work was the application of untargeted metabolomics approach with the novel combination of ion mobility separation coupled to high resolution mass spectrometry (IMS-HRMS) and study the advantages that this technique can bring to the area of dietary biomarker discovery, with the specific case of biomarkers associated with orange consumption (*Citrus reticulata*) in plasma samples taken during an acute intervention study (consisting of a randomized, controlled crossover trial in healthy individuals). A total of six markers of acute orange consumption, including betonicines and conjugated flavonoids, were identified with the experimental data and previous literature, demonstrating the advantages of ion mobility in the identification of dietary biomarkers and the benefits that an additional structural descriptor, as the collision cross section value (CCS), can provide in this area.

Keywords: orange intake; metabolomics; ion mobility; biomarkers

1. Introduction

Currently, one of the main limitations of nutritional epidemiology is the difficulty involved in measuring dietary intake [1]. In observational studies carried out on a large number of participants, the most commonly applied tools for estimating dietary intake are mainly based on self-reporting, including food frequency questionnaires (FFQs) for the assessment of regular consumption (usually one-year), or 24-h recalls for one-day assessment. However, such methodologies for data collection may contain substantial recall bias and other systematic or random errors that may have a large effect on the estimated food intake and, furthermore, on subsequent associations between food intake and the diseases studied [2,3]. Although in recent years, there has been an improvement in increasing the

Nutrients 2020, 12, 1916; doi: 10.3390/nu12071916

Ultra-performance liquid chromatography-ion mobility separation-quadrupole time-of-flight MS (UHPLC-IMS-QTOF MS) metabolomics for short-term biomarker discovery of orange intake: A randomized, controlled crossover study

by Leticia Lacalle-Bergeron[1], Tania Portolés[1], Francisco J. López[1], Juan Vicente Sancho[1], Carolina Ortega-Azorín[2,3], Eva M. Asensio[2,3], Oscar Coltell[3,4] and Dolores Corella[2,3]*

[1] Research Institute for Pesticides and Water (IUPA), Universitat Jaume I, 12071 Castellón, Spain

[2] Department of Preventive Medicine and Public Health, School of Medicine, University of Valencia, 46010 Valencia, Spain

[3] CIBER Fisiopatología de la Obesidad y Nutrición (CIBEROBN), Instituto de Salud Carlos III, 28029 Madrid, Spain

[4] Department of Computer Languages and Systems, Universitat Jaume I, 12071 Castellón, Spain

Abstract

A major problem with dietary assessments is their subjective nature. Untargeted metabolomics and new technologies can shed light on this issue and provide a more complete picture of dietary intake by measuring the profile of metabolites in biological samples. Oranges are one of the most consumed fruits in the world, and therefore one of the most studied for their properties. The aim of this work was the application of untargeted metabolomics approach with the novel combination of ion mobility separation coupled to high resolution mass spectrometry (IMS-HRMS) and study the advantages that this technique can bring to the area of dietary biomarker discovery, with the specific case of biomarkers associated with orange consumption (*Citrus reticulata*) in plasma samples taken during an acute intervention study (consisting of a randomized, controlled crossover trial in healthy individuals). A total of six markers of acute orange consumption, including betonicines and conjugated flavonoids, were identified with the experimental data and previous literature, demonstrating the advantages of ion mobility in the identification of dietary biomarkers and the benefits that an additional structural descriptor, as the collision cross section value (CCS), can provide in this area.

Keywords

Orange intake; Metabolomics; Ion mobility; Biomarkers

1. Introduction

Currently, one of the main limitations of nutritional epidemiology is the difficulty involved in measuring dietary intake [1]. In observational studies carried out on a large number of participants, the most commonly applied tools for estimating dietary intake are mainly based on self-reporting, including food frequency questionnaires (FFQs) for the assessment of regular consumption (usually one-year), or 24-h recalls for one-day assessment. However, such methodologies for data collection may contain substantial recall bias and other systematic or random errors that may have a large effect on the estimated food intake and, furthermore, on subsequent associations between food intake and the diseases studied [2,3]. Although in recent years, there has been an improvement in increasing the validity and precision of food questionnaires thanks to the use of new information technologies [4,5], these instruments are still biased and additional information based on objective intake biomarkers is needed [6,7]. Biomarkers of food intake are promising tools to provide more objective food consumption measurements [6-9]. Therefore, a major challenge nowadays for nutritional epidemiology and precision nutrition is to identify novel and valid biomarkers [10]. Metabolomics has opened up new opportunities for food intake biomarker discovery through metabolic profiling of biological samples (plasma, urine, etc.), following the intake of specific foods, meals, or diets [10,11]. Nevertheless, to date, there are very few validated biomarkers of food intake, and more research is urgently needed. In a consensus paper, Dragsted et al. [10] outlined an optimal and reproducible validation process to systematically and critically assess the validity of candidate food intake biomarkers. A consensus-based procedure was used to provide and evaluate a set of the most important criteria for systematic validation of these biomarkers [10,12]. Similar conclusions were reached in a National Institute of Health (NIH) organised workshop on “Omics Approaches to Nutritional Biomarkers” in Bethesda, attended by researchers from the United States, Canada, and several European countries [13]. According to these recommendations, acute interventions trials where participants consume specific amounts of a test food in a single meal are the best approach as the first step in dietary biomarker discovery [10-13]. In this approach, baseline

and postprandial biological samples are collected for analysis and for identifying potential biomarkers. For these dietary interventions, the use of crossover studies is an efficient design having less bias and requiring less sample size than randomised parallel group intervention trials [14]. In crossover studies, participants receive all interventions in consecutive periods, which means that the influence of individual variation is minimised. The metabolites are measured after every period. In these studies, the use of a single food, facilitates a more specific investigation of biomarkers, since the number of potential metabolites is simplified in comparison with a full dietary pattern. Despite this simplification, identifying metabolites for dietary intake is not always easy, even if only one food is administered [15]. Small, short-term feeding studies that yield candidate biomarkers may be followed by validation and testing of biomarker performance in large cohorts, repeating the process with required corrections, until an ideal biomarker performance is achieved. Currently, oranges are one of the foods for which it has been shown that there are good metabolomic biomarkers of intake [16-18]. Specifically, an excellent example of such a biomarker for citrus fruit intake is proline betaine (N,N-dimethyl-L-proline), also known as stachydrine [16-20]. This biomarker was first discovered in a small crossover trial [19], and later validated in several intervention and cohort studies [16,18,20-24]. Despite the high statistically significant associations found between citrus intake and proline betaine levels (both in plasma and/or in urine), correlation coefficients ranged between 0.4 and 0.8 depending on the study, suggesting that other additional biomarkers for citrus intake can add more information in precision nutrition. Some studies reported additional biomarkers for citrus fruit intake including trans-4-hydroxy-L-proline betaine (betonicine), N-methyl-L-proline and synephrine, among others [17,25-30]. However, with improvement in techniques and instruments, more biomarkers can be discovered. Moreover, it has been suggested that, whereas proline betaine is a good biomarker for the intake of any citrus product, synephrine can be used as a reliable additional biomarker with high specificity for orange intake [30].

Aside from specific compounds, in metabolomics the analytical platform used as well as the targeted or untargeted approaches, may have a

role to play in identifying metabolites [31], each of these having a number of advantages and disadvantages [31]. Food is usually a complex and diverse mixture of components that could present a metabolic potential or not. To analyse these compounds, there is a need for robust, efficient and sensitive methodologies along with powerful analytical technologies [32,33]. Metabolomic fingerprinting is the global screening approach (untargeted method) used for low molecular mass compounds (<1500 Da), for classifying the samples based on the metabolic profile or "fingerprints" that change in response to external (environmental, diet…) or internal (genetic, disease…) perturbations with the final goal of identifying or discriminating between metabolites [15,33,34].

Due to the large amount of compounds that are joined in the metabolome, with a large variety of physicochemical properties, in recent years the combination of separation techniques such as gas or liquid chromatography (GC and LC respectively) with high resolution accurate mass spectrometry (HRMS) are appearing to be less expensive and more sensitive alternative approaches than nuclear magnetic resonance (NMR) [35], which has been the previous technique of choice [36]. Because of the physical properties of most biofluids, GC is not usually chosen in this field because the analytes (usually polar and not very volatile) often requires derivatisation to improve thermal stability for GC, requiring more laborious sample treatment [37,38]. Due to the aqueous composition of biofluids (such as blood, urine, sweat), LC has been widely used for metabolomics in bioanalytical studies. Furthermore, a wider range of separation mechanisms due to the variety of stationary phases is available, so that complex sample preparation is not usually required, and shorter separation run times are involved [39].

Therefore, hyphenation LC-HRMS is widely employed in untargeted metabolomic approaches. Moreover, the recent introduction of ion mobility separation (IMS) coupled to HRMS Q-TOF provides additional structural information, so achieving a better characterisation of biomarkers [40-42]. IMS is a gas-phase separation technique based on the time that an ion takes to cross a drift tube filled with an inert gas under low-field conditions. The drift time (DT) of each ion is dependent on its individual size, shape and charge [43,44]. DT can be directly converted into a collision cross section

value (CCS, Å^2), which represents an additional separation dimension based on the compound chemical structure and 3-dimensional conformation [44,45]. The LC-IMS-HRMS combination provides an extra separation dimension through CCS to the retention time (RT), accurate mass (*m/z*) and intensity, thus obtaining four-dimensional matrix data.

Therefore, the main aim of this study was to perform an exploratory untargeted metabolomic study using UHPLC-IMS-HRMS to find out short-term plasma biomarkers of orange consumption, as well as to explore the promising improvements that IMS in combination with LC-HRMS could achieve in the untargeted metabolomic field in an acute crossover intervention trial using multivariate analysis (PCA, PLS-DA and OPLS-DA) to highlight the most relevant short-term markers of orange intake.

2. Materials and Methods

2.1. Subjects and Study Design

We carried out a randomised, controlled crossover study on 30 healthy subjects (aged 25.0 ± 2.8 years, 8 females and 22 males), recruited in Valencia, Spain. After a minimum of 8 h fasting, participants were randomly allocated to eat 500 g of peeled oranges (*Citrus reticulata, Clementine variety*) or an isocaloric (same energy as the oranges) solution of sucrose in water. No other food was allowed for 4 h (a standard time-point for acute human studies using plasma samples). At the start and after 4 h, blood samples were taken. Subjects were randomised to start either in the intervention or the control groups in order to prevent the influence of period-effects. In addition, a wash-out period was considered. After a one-week break the two groups swapped over and the corresponding intervention was undertaken. The crossover study was registered as ISRCTN17330010. Anthropometric, clinical, lifestyle and biochemical data were obtained by standardised techniques and questionnaires as previously described [46]. Plasma was obtained and stored at −80 °C for subsequent metabolomics determinations. Subjects provided written informed consent and the study protocol and procedures were approved according to the ethical standards of the Helsinki Declaration and by the Human Research Ethics Committee of the University of Valencia, Valencia (reference number: H1425917369903).

2.2. Chemicals and Reagents

LC-MS grade methanol (MeOH) and LC-MS grade acetonitrile (ACN) were purchased from Sharlab (Barcelona, Spain), as well as Formic acid (HCOOH) eluent additive for LC-MS and reagent grade ammonium acetate (NH_4Ac). Milli-Q water purification system (Millipore Ltd., Bedford, MA, USA) was used to obtain HPLC-grade water. Leucine-enkephalin (mass-axis calibration) was purchased from Sigma-Aldrich (Saint Louis, MO, USA).

2.3. Sample Treatment

Prior to analysis, the blood plasma samples were thawed at room temperature and vortexed. A total of 400 µL of ACN was added to 100 µL of sample. The supernatant was collected after centrifuging 12,000×rpm and a radius of 5.5 cm, 8855 g (RCF), for 10 min at 4 °C, and divided into 3 aliquots: two vials of 100 µL were stored at −30 °C, one vial of 200 µL was stored at −80 °C. Moreover, 20 µL of each sample were pooled and mixed to generate the Quality Control sample (QC). This is assumed to provide a representative average sample formed by a pool equivalent aliquot of all final sample extracts.

In order to compensate/reduce instrumental drift, samples were randomly injected into the UHPLC-IMS- QTOF MS system. QC was used for both column stabilization purposes (by 10 QC injections at the beginning of each sample batch) and, following the injection of every 10 samples, to control possible instrumental drift throughout the sequence.

2.4. Instrumentation

Ultra-high performance liquid chromatography (UHPLC) with a ACQUITY UHPLC I-Class system (Waters, Milford, MA, USA) was coupled to a VION® IMS QTof (Waters, Manchester, UK), ion mobility hybrid Quadrupole Time-of-Flight (IMS-QTOF) High Resolution Mass Spectrometer (UHPLC-IMS-HRMS) using a electrospray ionization interface operating in both positive (ESI+) and negative (ESI−) mode. Equipment control and data acquisition and processing were performed using UNIFI software (V.1.8.2, Waters, Manchester, UK).

 Nutrients 2020, 12,1916; doi: 10.3390/nu12071916

2.4.1. UHPLC Conditions

Two LC separation procedures were employed to cover a wider range of compound polarities. To separate the medium to nonpolar molecules, Reversed Phase Liquid chromatography (RP) was used with a CORTECS® C18 fused-core 2.7 µm particle size analytical column 100 × 2.1 mm (Waters), whereas a CORTECS® HILIC fused-core 2.7 µm particle size analytical column 100 × 2.1 mm (Waters) was employed for Hydrophilic Interaction Liquid Chromatography (HILIC) to separate the polar molecules. 0.3 mL/min flow rate, 40 °C column oven temperature and 1 µL sample injection volume were selected for both types of liquid chromatography and both ionization modes.

The RP-LC gradient elution was performed using mobile phases A = H_2O and B = MeOH, both with 0.01 % of HCOOH. The gradient changed from 10 % B at 0 min to 90 % B at 14 min, 90 % B at 16 min, and 10 % B at 16.01 min, with a total run time of 18 min. The same gradient was employed for both ESI+ and ESI− ionization modes.

For HILIC separation, mobile phases A = ACN:H_2O (95:5, *v/v*) and B = H_2O, both with 0.01 % HCOOH and 10 mM NH_4Ac, were employed. The gradient started with 2 % B until 1.50 min, 15 % B at 2 min, 50 % B at 6 min, 60 % B at 7.50 min and finally 2 % B at 7.51 min, with a total run time of 10 min. This gradient was the same for both ESI+ and ESI− ionization modes.

2.4.2. IMS-QTOF Set Up

The capillary voltage was set at 0.7 kV and 2.00 kV for positive (ESI+) and negative (ESI−) electrospray ionization mode, respectively. Nitrogen was used as the desolvation gas, nebulizing gas and collision gas. The source temperature was set to 120 °C and desolvation gas to 550 °C with a flow rate of 1000 L/h. The mass spectrometer was operated in ion mobility ($HDMS^E$) mode for acquisition in both polarities. In $HDMS^E$ experiments, two acquisition functions were acquired simultaneously over an *m/z* range of 50–1000 Da and a scan time of 0.3 s. Low Energy function (LE), with a fixed collision energy of 6 eV, and High Energy function (HE) with a collision energy ramp from 28 to 56 eV was set.

Calibrations of mass axis and DT were performed monthly with the "Major Mix IMS/T of calibration kit" supplied by the vendor (Waters), infused at a flow rate of 20 μL/min for both positive and negative mass calibrations as well as CCS calibration.

For automated mass measurement, a Leucine-Enkephalin solution (100 ppb) in ACN:H_2O (50:50, *v/v*) at 0.01 % HCOOH was pumped at 20 μL/min through the lock-spray needle and measured every 5 min (ensuring a measurement at the beginning, in the middle and at the end of the chromatogram), with a scan time of 0.3 s. The (de)protonated molecule of leucine-enkephalin, at *m/z* 556.27658 in ESI+ and *m/z* 554.26202 in ESI– was employed to recalibrate the mass axis and ensure the robust accurate mass measurement along runs.

2.5. Data Processing and Statistical Analysis

The VION instrument data (*.uep*, UNIFI, Waters) were imported to Progenesis QI (V.2.3, NonLinear Dynamics, Waters, Newcastle, UK). Then, the software automatically performs the retention time alignment, with the QC samples as reference (except for the first 9 QC injections used for column stabilization); the 4D peak picking (based on the intensity, *m/z*, retention time and DT) and response normalisation. The peak picking conditions were set as follows: all runs, limits (automatic), sensitivity (automatic, level 2), chromatographic peak width (minimum peak width of 0.1 min), and retention time limits (0.3 to 17 min and 0.3 to 9 min, for RP and HILIC respectively). To apply the deconvolution tool, the selected adducts ions forms $[M+H]^+$, $[M-H_2O+H]^+$, $[M+Na]^+$ and $[M+K]^+$ were selected for positive ionization analysis and $[M-H]^-$, $[M-H_2O-H]^-$, $[M+Cl]^-$, $[M+FA-H]^-$ for negative ionization analysis. Samples were divided into 4 groups (Orange t = 0 h, Orange t = 4 h, Isocaloric Beverage (IB) t = 0 h, IB t = 4 h) in the "*Experimental Design Setup*", following the "*Within-subject Design*" comparison, where not only are the groups each sample belongs to specified to the software, but also the subject from which it comes. The software will then perform a repeated measures ANOVA, to reduce the individual differences.

The processed data were then directly exported to EZinfo (V.3.0, Umetrics, Sweden) for multivariate statistical analysis. First, Principal Component Analysis (PCA), an unsupervised analysis, was applied to ensure the correct grouping of the QC samples in the centre of the plot after normalisation and the absence of outliers. Then, Partial Least Square–Discriminant Analysis (PLS-DA) was conducted to maximize the separation between the groups. Ultimately, an Orthogonal PLS-DA (OPLS-DA) was carried out to highlight the most robust markers (threshold $p(corr) \geq |0.6|$).

2.6. Elucidation Workflow

The accurate mass and retention time of the most significant markers from the OPLS-DA were obtained from the feature table and they were checked and sought in the raw data (by UNIFI Platform). The compounds were tentatively elucidated with the aid of an elemental composition calculator (UNIFI, V1.8.2, Waters). The DT filtered HE spectra of the target feature was searched in reliable mass spectra data bases (HMDB, MetLin, MassBan) or were compared to in-silico fragmentation spectra (MetFrag) with subsequent searches through general chemical data bases such as Chemspider and PubChem.

The final identity could only be confirmed by comparing the retention time, fragmentation and CCS with a commercially available standard. When not available, CCS values were predicted by means of our CCS prediction model [45] aimed at providing additional identification confidence.

3. Results

3.1. Participants and Experimental Setup

Table 1 shows the general characteristics of the study participants. A total of 30 healthy subjects (97 % non-smokers) were analysed in the intervention study and all of them completed both the dietary intervention with oranges and the control group (isocaloric intervention with sucrose) at baseline and after 4 h. Plasma samples were obtained twice at baseline and after 4 h and analysed for metabolomics biomarkers of orange intake.

Regarding sample treatment, the only step taken was deproteinisation with ACN to eliminate the macromolecules (nucleic acids and proteins)

present in the plasma and, thereby, avoid possible interferences in the metabolomic analysis and so focus on the low-weight molecules (metabolites). The sample treatment must be as low selective as possible to cover a larger range of compounds. The supernatant was directly injected in the UHPLC-IMS-QTOF MS system. The samples extracts were injected in randomised order to avoid bias in the methodology and QC was injected every 10 samples for normalisation and instrumental drift control.

Table 1. *Demographic, anthropometric and clinical characteristics of the participants by sex.*

Characteristics	Total (*n*=30)	Men (*n*=22)	Women (*n*=8)
Age (years)	25.0 ± 0.5	25.4 ± 0.6	23.8 ± 0.9
BMI (Kg/m^2)	25.0 ± 1.0	25.8 ± 1.3	22.6 ± 1.1
SBP (mm Hg)	124 ± 3	129 ± 3	111 ± 3
DBP (mm Hg)	74 ± 2	75 ± 2	74 ± 2
Total cholesterol (mg/dL)	186.7 ± 5.9	191.4 ± 6.8	172.3 ± 11.6
LDL-C (mg/dL)	121.5 ± 4.6	126.8 ± 5.0	104.9 ± 8.8
HDL-C (mg/dL)	57.5 ± 2.7	53.6 ± 2.7	68.2 ± 5.4
Triglycerides (mg/dL)	75.4 ± 7.5	83.6 ± 9.7	52.8 ± 3.4
Fasting glucose (mg/dL)	86.7 ± 1.1	88.2 ± 1.3	82.7 ± 1.6

Values are mean ± SE for continuous variables. BMI indicates body mass index; SBP indicates Systolic Blood Pressure, DBP indicates Diastolic Blood Pressure; LDL-C indicates Low-Density Lipoprotein cholesterol; HDL-C indicates High-Density Lipoprotein cholesterol.

Owing to the aqueous nature of the plasma samples, LC was the most convenient chromatographic separation technique. Furthermore, due to the high sensitivity of the instrument employed, only a low volume injection was possible (1 µL). This allowed the samples extracts (ACN:H_2O, 80:20 *v/v*) to be directly injected not only in HILIC, where the initial chromatographic conditions were predominantly in the organic phase; but also in RP, without the need to evaporate and redissolve in a more suitable solvent mixture (predominantly aqueous medium). RP and HILIC were employed to detect as many compounds as possible, in positive (pos) and negative (neg) ionization modes. RP (C18 column) is more suited to nonpolar hydrophobic molecules and HILIC (silica-based column) for polar hydrophilic compounds. Finally, we obtained four data sets for the samples under study (RP pos, RP neg, HILIC pos and HILIC neg).

All data were acquired in HDMS^E, where, apart from the DT (ms) measurement by ion mobility, LE and HE spectra were also acquired simultaneously, obtaining precursor ion information and full-scan accurate mass fragmentation information, respectively. Therefore, at the end of the analysis, a four-dimensional data matrix was obtained, which allowed each feature to be characterised by means of the retention time from chromatographic separation, the intensity, the CCS calculated from the DT together with the accurate mass and fragmentation spectra. Moreover, the cleaned HE spectra afforded by the DT separation enhanced the structural elucidation since it only shows the fragments that have been generated from a "precursor" ion with a given DT. This is because the ion mobility separation prior to the fragmentation ensures that the precursor ion shares the same DT as its fragments on the three-dimensional plots, allowing one to align the feature and its fragments and reduce the interference of co-eluting components, thus rendering a HE spectra of MS/MS quality.

3.2. Data Processing and Analysis

The four data sets (RP pos, RP neg, HILIC pos and HILIC neg) were acquired with UNIFI software (Waters, UK) and the raw data were exported to **.uep* format (unifi export package). To the best of our knowledge, there is no program but Progenesis QI (NonLlinear Dynamics, Waters, Uk) able to interpret this data format and process four-dimensional data (RT, *m/z*, area and CCS) for -omics purposes. The processing workflow in Progenesis QI starts with data import, followed by retention time alignment, peak picking and normalisation. The alignment score values for all runs were higher than 85 % and normalisation was performed using all compounds. The peak picking resulted in the detection of 6951, 6238, 5283 and 4479 ions in RP pos, RP neg, HILIC pos and HILIC neg, respectively. According to the adduct ions specified, the deconvolution tool allowed us to group the features coming from the same compound and annotate them under a single label (*xx.xx_yyy.yyyyn*, *xx.xx* being the retention time in minutes and *yyy.yyyy* the exact neutral mass when more than one adduct ion is found for the same compound; or *xx.xx_zzz.zzzzzm/z*, *zzz.zzzz* being the exact mass of the single ion found).

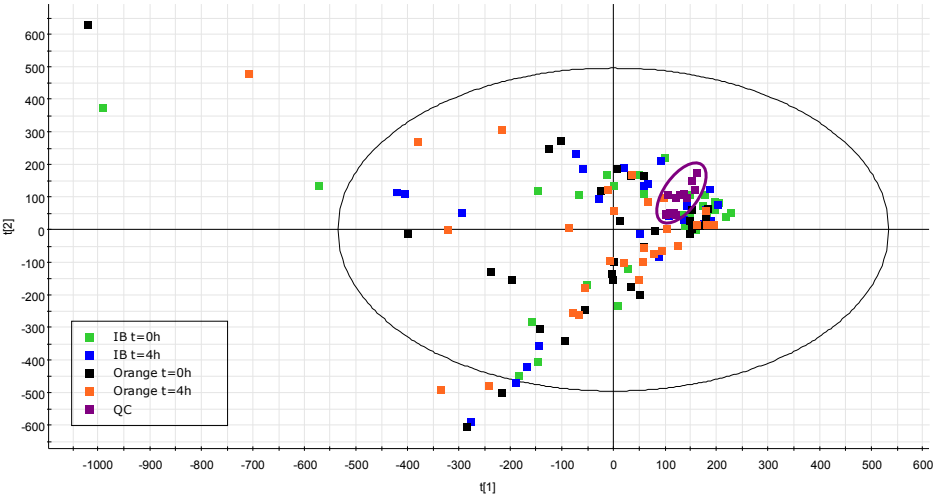


Figure 1. *PCA score plot component 1 vs. component 2 of Hydrophilic Interaction Liquid Chromatography (HILIC) in positive ionization mode, explaining the 20.6 % and 17.7 % of the variance, respectively. The purple points (14 QC samples) are grouped and centred in the plot. Four plasma samples were obtained from each of the 30 participants obtaining 30 samples per group. A total of 120 plasma samples were analysed. The samples obtained at t = 0 h and t = 4 h after the intake of an isocaloric beverage (IB) and the samples obtained at t = 0 h and t = 4 h after orange intake are coloured in green, blue, black and orange, respectively.*

PCA was applied to each data set, so providing a non-supervised exploratory visualisation of the results to detect possible outliers and to ensure that the differences between the groups were not coming from the instrumental drift over time or error in the data treatment. For this purpose, QC injections were injected at the beginning of the batch for column conditioning and after every 10 samples. The QCs (n = 14 per data set) should be grouped in the centre of the PCA plots, as they behave as an "average sample" (pool of all the samples analysed) and thus demonstrate the correct acquisition of the data, as well as that the differences between the groups are not caused by instrumental processing. **Figure 1** shows the PCA of the HILIC pos data set, where non-inherent differentiation between the group of samples can be observed. Although, the grouping of the QC samples near to the centre of the plot proves the proper performance of the analytical system along the run. PCA plots of the data sets HILIC neg, RP pos and RP neg are shown in **Figures S1–S3** respectively.

As this was a crossover trial, control samples of each participant were collected and a "Within-subject Design" comparison performed. In this type of experimental design, where samples of each participant are taken in different conditions, not only must the time the sample was taken be specified, but also which subject it came from. In these cases, standard ANOVA is not appropriate because data violates the independence ANOVA assumption and a repeated measures ANOVA should be applied. The data were filtered by means of repeated measures ANOVA *p-value* ≤ 0.05 to reduce the individual differences and focus on the potential markers. Thus, the data matrices were reduced to 262, 302, 298 and 149 relevant features in RP pos, RP neg, HILIC pos and HILIC neg, respectively. These reduced data sets were joined into a single file (1011 features data matrix).

The next step was to apply a supervised multivariate statistical method by means of PLS-DA modelling to highlight the differences between the preselected groups. This analysis was applied to the compendium of the reduced data sets per ANOVA. **Figure 2** shows a remarkable discrimination between samples taken after orange consumption and those obtained in fasting or after isocaloric beverage intake (with 38.3 % of the variance explained with two components).

Moreover, these last three groups are mixed and behave in a similar way. Thus, an OPLS-DA was applied contrasting the samples collected 4 h after orange intake vs. the other samples (**Figure S4** shows the scatter plot of the OPLS-DA). From OPLS-DA an S-Plot was generated (**Figure 3**), each feature being assigned a number between −1 and 1 according to their discrimination power between the two groups called *p(corr)*. Consequently, the most relevant features in the discrimination are in the extreme parts of the plots: those with a higher presence in the orange intake sample have a *p(corr)* near to 1 and those significant for the other group have a *p(corr)* near −1. To ensure the validity of these features as possible markers, a PCA analysis was conducted with only the 43 highlighted features in the S-Plot as variables. As can be noticed in **Figure S5**, the samples obtained before and after the isocaloric beverage and the samples before the orange consumption are clustered and completely separated from the samples obtained 4 h after the orange intake. **Figure 3** shows that, from the total of 1011 features, only

43 from all data sets with a repeated measures ANOVA p-value ≤ 0.05 were highlighted as discriminatory between the samples after orange consumption and the group of composite samples without orange intake with a *p(corr)* ≥ |0.6|. Nevertheless, the list of possible markers could be reduced still further to seven features for various reasons. For example, two features higher in the orange sample with the same exact mass and the same CCS (127.18 Å^2) but different retention time (HIpos_5.02_143.0941n and RPpos_0.64_143.0947n, with *p(corr)* = 0.96 and 0.92, respectively) corresponded to the same compound, but detected under HILIC and RP separation. In other cases, some types of ions (such as dimers) had not been specified for the Progenesis QI deconvolution step, and therefore these adducts could appear as independent features in the statistics.

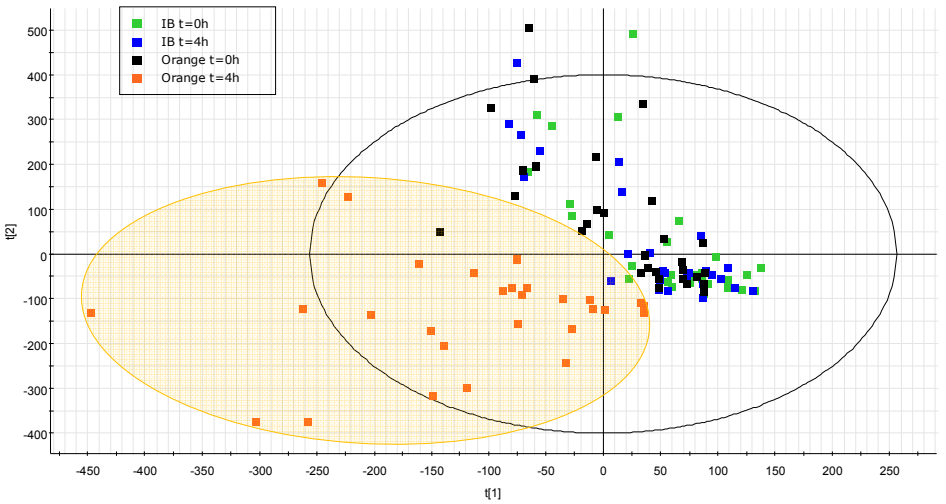


Figure 2. *PLS-DA score plot based on the features with repeated measures ANOVA p-value ≤ 0.05. A 2D vision component 1 vs. component 2 with 24.5 % and 13.8 % of the variance explained, respectively. The samples obtained at t = 0 h and t = 4 h after the intake of an isocaloric beverage (IB) and the samples obtained at t = 0 h and t = 4 h after orange intake are coloured in green, blue, black and orange, respectively.*

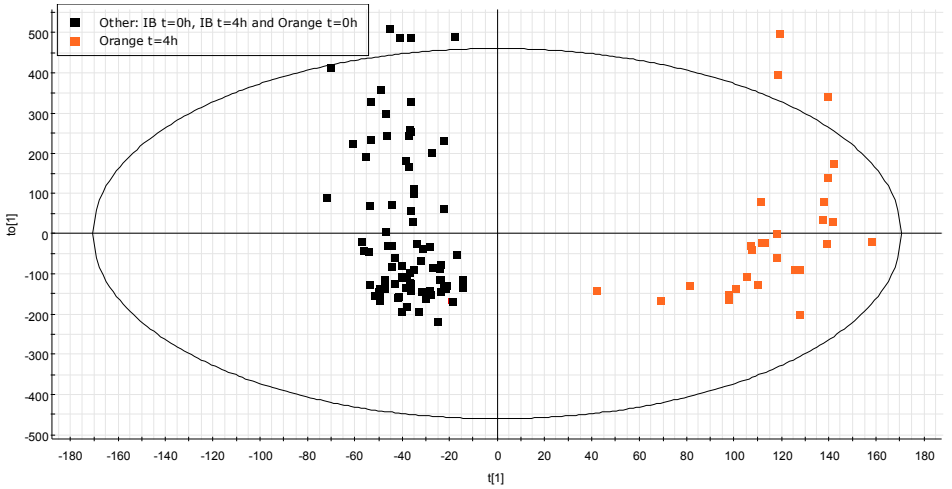


Figure S4. *OPLS-DA score plot based on the features with repeated measures ANOVA p-value ≤ 0.05, where the 94 % of the total variance is explained. The samples obtained at t = 0 h and t = 4 h after the intake of an isocaloric beverage (IB) and the samples obtained at t = 0 h are marked as "others" group and coloured in black; while samples obtained t = 4 h after orange intake are coloured in orange.*

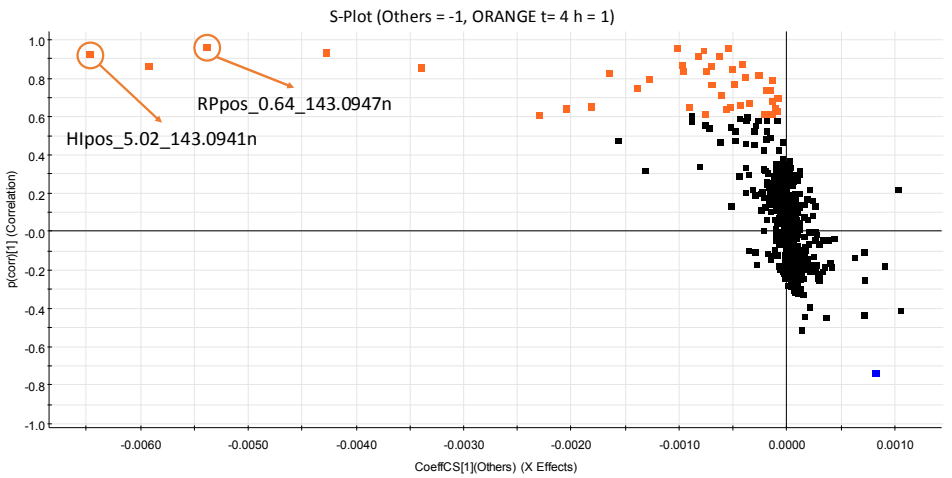


Figure 3. *S-Plot from the OPLS-DA where features with repeated measures ANOVA p-value ≤ 0.05 are represented. Tentative markers with a p(corr) higher than 0.6 are highlighted in orange and lower than –0.6 in blue.*

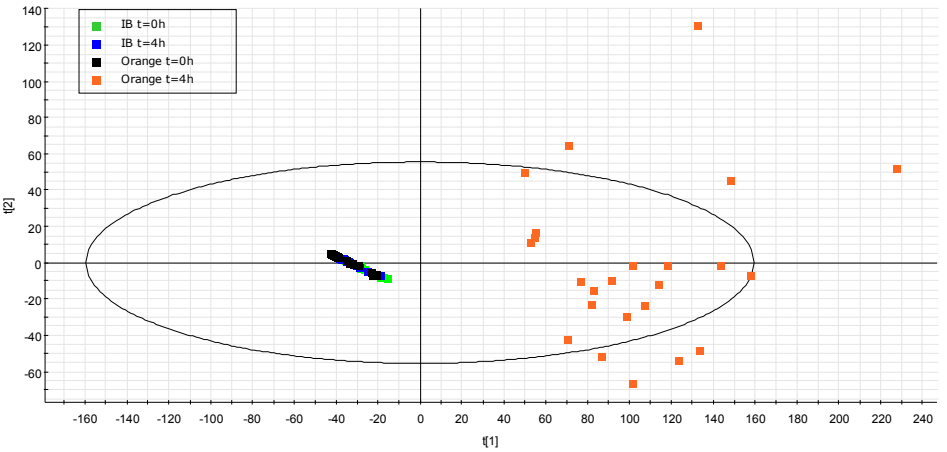


Figure S5. *PCA score plot component 1 vs component 2 obtained with the 43 features highlighted in the OPLS-DA with a p(corr) ≥ |0.6|, explaining the 79.3 % and 9.6 % of the variance, respectively. The samples obtained at t = 0 h and t = 4 h after the intake of an isocaloric beverage (IB) and the samples obtained at t = 0 h and t = 4 h after orange intake are coloured in green, blue, black and orange, respectively.*

3.3. Elucidation Process

A total of 7 statistically representative compounds were tentatively identified (**Figure 4**, **Table 2**) as early markers of orange consumption in plasma. In order to accomplish the elucidation, the $HDMS^E$ spectra were studied to determine the candidate identity based on mass accuracy and both parents and fragment ions.

Marker 1 elucidation workflow was selected as an illustrative example of the elucidation process. **Figure 5** shows the differences obtained in the mass spectra when it is filtered or not by the DT of the parent ion (4.97 ± 0.21 ms). The ion mobility separation prior to the fragmentation implies that both parent and fragment ions will have the same recorded DT. Hence, it is possible to filter the target ion by DT and obtain cleaner spectra without interfering ions appearing at the same retention time for the LE spectrum and without co-elutant fragments in the HE spectrum. The possibility of visualising only "product" ions derived from a specific "precursor" ion, allows us to obtain a quasi-MS^2 spectrum, without the need for reinjecting the samples.

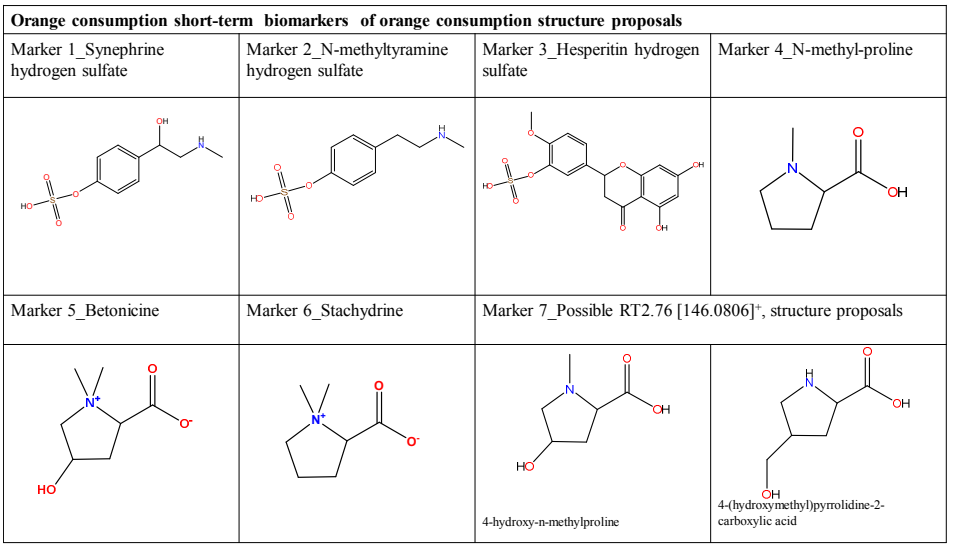


Figure 4. *Tentative structures and identity of the makers elucidated.*

However, information was also obtained from the unfiltered spectra. Since the adduct ions are formed before ion mobility separation, their DT could be different, and therefore they would not be shown in the DT filtered spectrum. In the case of Marker 1 (HIPOS_2.17_247.0510n, *p(corr)* = 0.64), as can be seen in **Figure 5A**, the dimers $[2M+H]^+$ and $[2M+Na]^+$ are also present, as dimers were not selected as possible "adduct" ions in the Progenesis QI deconvolution, but were assigned as independent features. These features were also highlighted as discriminant in the OPLS-DA with a *p(corr)* > 0.6.

Marker 1 was also found in HILIC neg analysis (HINEG_2.22_246.0434*m/z*) with different ion forms. The extracted ion chromatograms for the ions and dimers founded corresponding to this marker are shown in **Figure S6**, and all of them are highlighted as markers. The most likely elemental composition for this marker was found to be $C_9H_{17}NO_5S$ (error: −0.4 and −0.07 mDa in positive and negative HILIC analysis, respectively). The mean retention time across the samples was 2.2 min and CCS value 158.27 Å^2 and 155.40 Å^2 for protonated and deprotonated ion, respectively. **Figure 6** shows the LE and HE spectra obtained in HILIC pos and HILIC neg analysis, filtered by the DT of the protonated and deprotonated ion, respectively.

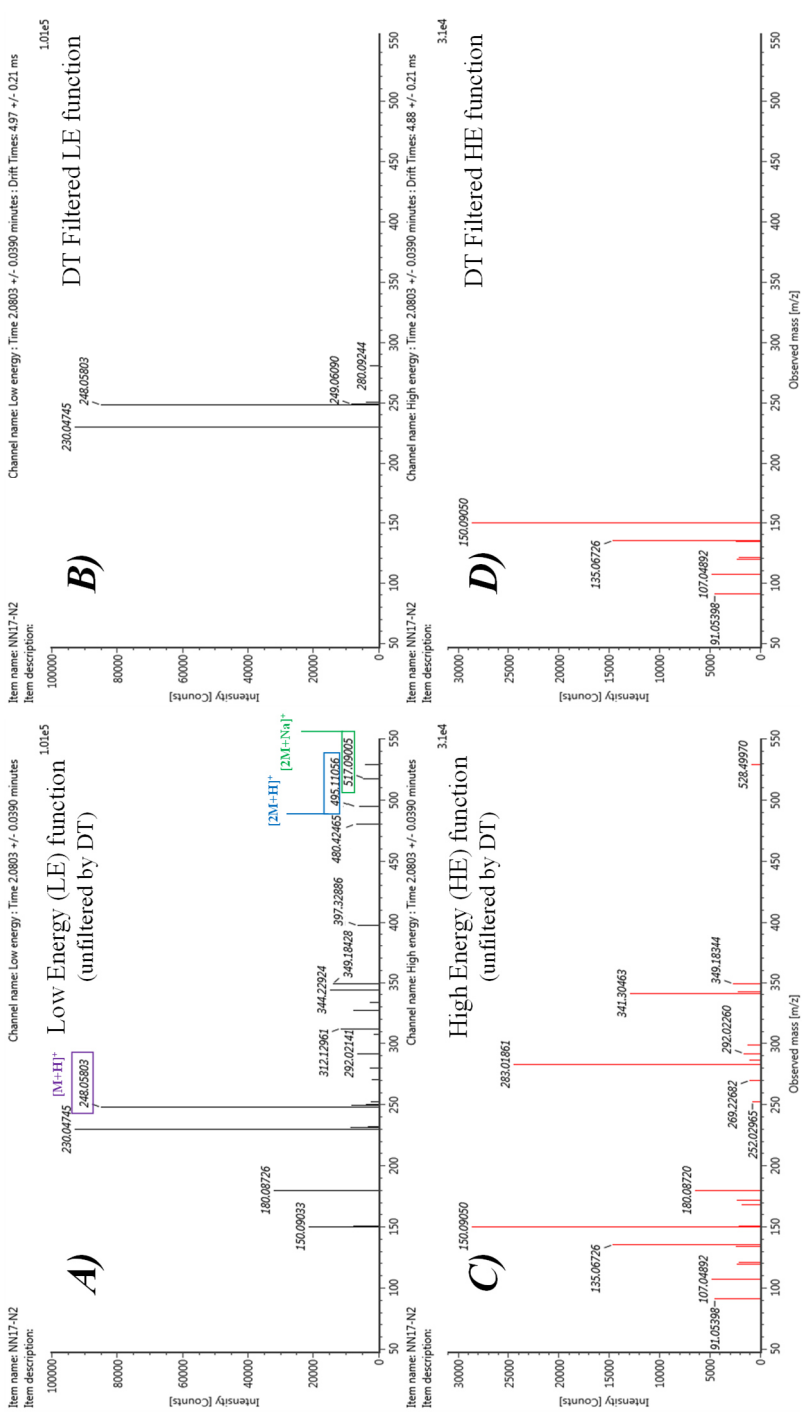


Figure 5. *HDMS^E spectra from HILIC pos analysis for Marker 1 with and without DT filtering: (**A**) LE function (without DT filtering), (**B**) DT filtered LE function, (**C**) HE function (without DT filtering), and (**D**) DT filtered HE function.*

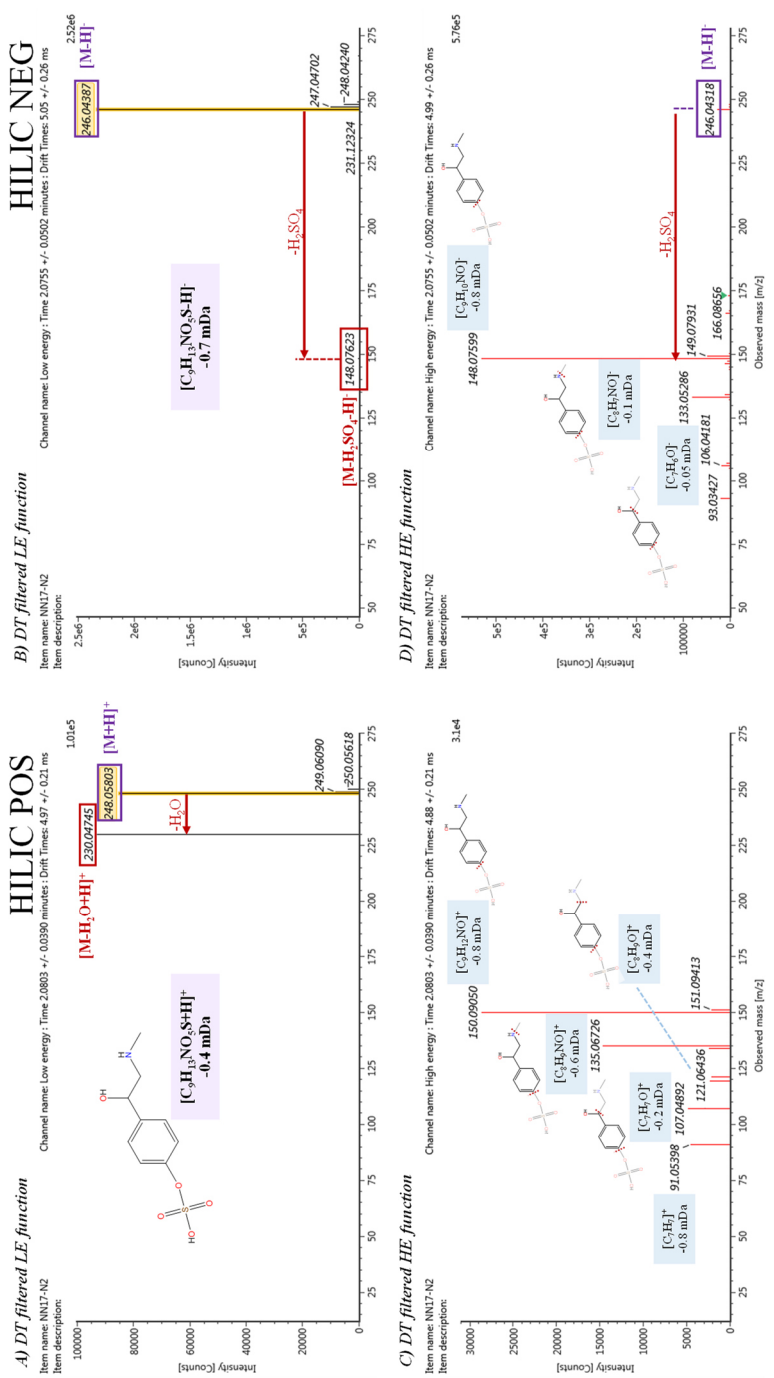


Figure 6. *DT filtered HDMS^E obtained for Marker 1: (**A**) LE function in HILIC pos analysis, (**B**) LE function in HILIC neg analysis, (**C**) HE function in HILIC pos analysis and (**D**) HE function in HILIC neg analysis. A proposed structure and fragmentation are shown.*

In the HILIC pos LE function, an in-source fragment *m/z* 230.0476 ($[C_9H_{12}NO_4S]^+$, error: −0.7 mDa) corresponding to a neutral loss of H_2O (not very specific fragment) can be seen. From the HILIC neg LE function is obtained a more specific in-source fragment *m/z* 148.0762 ($[C_9H_{10}NO]^-$, error: −0.7 mDa) corresponding to the neutral loss of H_2SO_4, a typical loss of phase II sulfate conjugated metabolites. This ion is also present in the HILIC neg HE spectrum as a product ion, with *m/z* 133.0529 ($[C_8H_7NO]^-$, error: −0.1 mDa) assigned as consecutive neutral loss of CH_3, followed by neutral loss of HCN with *m/z* 106.0418 ($[C_7H_6O]^-$, error: −0.05 mDa). These ions were also present in HILIC pos HE spectra in their corresponding positive adducts *m/z* 150.0905 ($[C_9H_{12}NO]^+$, error: −0.8 mDa), *m/z* 135.0673 ($[C_8H_9NO]^+$, error: −0.6 mDa) and *m/z* 107.0418 ($[C_7H_7O]^+$, error: −0.2 mDa), among other product ions. After the automated searching in databases with the UNIFI (Waters) tool, in-silico fragmentation tools and spectra comparison with data bases as HMDB if it is available, Synephrine hydrogen sulfate, phase II sulfate conjugated metabolite of synephrine, was the most likely tentative identification, and this match was supported by the fragmentation spectra.

Synephrine is a citrus compound and commonly present in dietary supplements (coming from citrus extracts) used in weight loss, and therefore its metabolism and toxicity has been extensively studied [47]. Moreover, it has previously been reported as a biomarker of citrus intake in urine, along with other metabolites of other flavonoids citrus compounds as Marker 2 and 3, tentatively identified as phase II sulfate conjugated metabolites of N-Methyltyramine and Hesperitin, respectively (**Table 2** and **Table 3**). These compounds have been previously studied due to their antioxidants and anti-inflammatory properties [48,49]. Moreover, hesperitin phase I/II conjugates have already been described as major metabolites of other citrus containing the flavonoids hesperidin and naringenin [50,51].

Table 2. *Compound list obtained from the untargeted metabolomic approach for determining acute biomarkers of orange consumption.*

No.	Compound	Elemental composition	P[corr]	Feature	Rt (min)	Experimental neutral mass (Da)	Theoretical neutral mass (Da)	Mass error (mDa/ppm)	CCS (Å²) de/protonated molecule *m/z*	Predicted CCS (Å²) protonated molecule *m/z*[1]	CCS delta error (%)	Adducts detected
1	Synephrime hydrogen sulfate	$C_9H_{13}NO_5S$	0.64	HIPOS_2.17_247.0510n	2.17	247.0510	247.0514	-0.4/-1.6	158.27	149.79	-5.36(1)	$[M+H]^+$ $[M-H2O+H]^+$ $[2M+H]^+$ $[2M+Na]^+$
			0.86	HINEG_2.22_246.0434m/z	2.22	247.0507	247.0514	-0.7/-2.8	155.40	---	---	$[M-H]^-$ $[(M-H+Na)-Cl]^-$ $[2M-H]^-$ $[2M-2H+Na]^-$
2	N-methyltyramine hydrogen sulfate	$C_9H_{13}NO_4S$	0.63	HIPOS_1.89_232.0629m/z	1.89	231.0556	231.0565	-0.9/-3.9	154.00	146.16	-5.09(1)	$[M+H]^+$
			0.82	HINEG_1.89_230.0483m/z	1.89	231.0556	231.0565	-0.9/-3.9	154.49	---	---	$[M-H]^-$
3	Hesperitin hydrogen sulfate	$C_{16}H_{14}O_8S$	0.61	HINEG_0.58_381.0279m/z	0.58	382.0352	382.0359	-0.7/-1.8	175.63	---	---	$[M-H]^-$
			0.57	HIPOS_0.57_383.0430m/z	0.57	382.0357	382.0359	-0.2/-0.52	172.50	186.02	-7.83(1)	$[M+H]^+$
4	N-methyl-proline	$C_6H_{11}NO_2$	0.84	HIPOS_4.73_130.0856m/z	4.73	129.0783	129.0790	-0.7/-5.4	127.17	124.4	-2.18	$[M+H]^+$
5	Betonicine	$C_7H_{13}NO_3$	0.93	HIPOS_4.79_159.0890n	4.79	159.0890	159.0895	-0.5/-3.1	129.02	130.25	+0.95	$[M+H]^+$ $[M+Na]^+$ $[M+K]^+$ $[2M+Na]^+$
			0.92	RPPOS_0.66_160.0970m/z	0.66	159.0897	159.0895	+0.2/+1.3	129.02	130.25	+0.95	$[M+H]^+$ $[M+Na]^+$
6	Stachydrine	$C_7H_{13}NO_2$	0.92	HIPOS_5.02_143.0941n	5.02	143.0941	143.0946	-0.5/-3.5	127.17	127.55	+0.30	$[M+H]^+$ $[M+Na]^+$ $[M+K]^+$ $[2M+H]^+$ $[2M+Na]^+$
			0.96	RPPOS_0.64_143.0947n	0.64	143.0947	143.0946	+0.1/+0.7	127.18	127.55	+0.29	$[M+H]^+$ $[M+Na]^+$
7	unknown	$C_6H_{11}NO_3$	0.65	HIPOS_2.76_146.0806m/z	2.76	145.0733	145.0739	-0.6/-4.1	123.81	127.11/127.03	+2.66/+2.60	$[M+H]^+$

(1)Sulfates and glucuronides error in the prediction, not published yet the correction

In the case of Marker 6, **Figure 7** shows the extracted ion chromatograms and HDMS^E spectra in HILIC pos analysis. This marker was also found in the dead volume in RP pos due to its high polarity, corresponding to the features HIpos_5.02_143.0941n and RPpos_0.64_143.0947n highlighted in **Figure 3**, among other features corresponding to dimers. Similar behaviour occurred in Marker 5. These two markers were the ones with higher *p(corr)* and abundancy. Therefore, they are the most differentiating markers present in the plasma samples 4 h after orange consumption (**Table 2**). These markers were tentatively elucidated as Betonicine and Stachydrine (proline betaine). These betaines have been extensively reported as a citrus intake biomarker in blood and urine (**Table 3**).

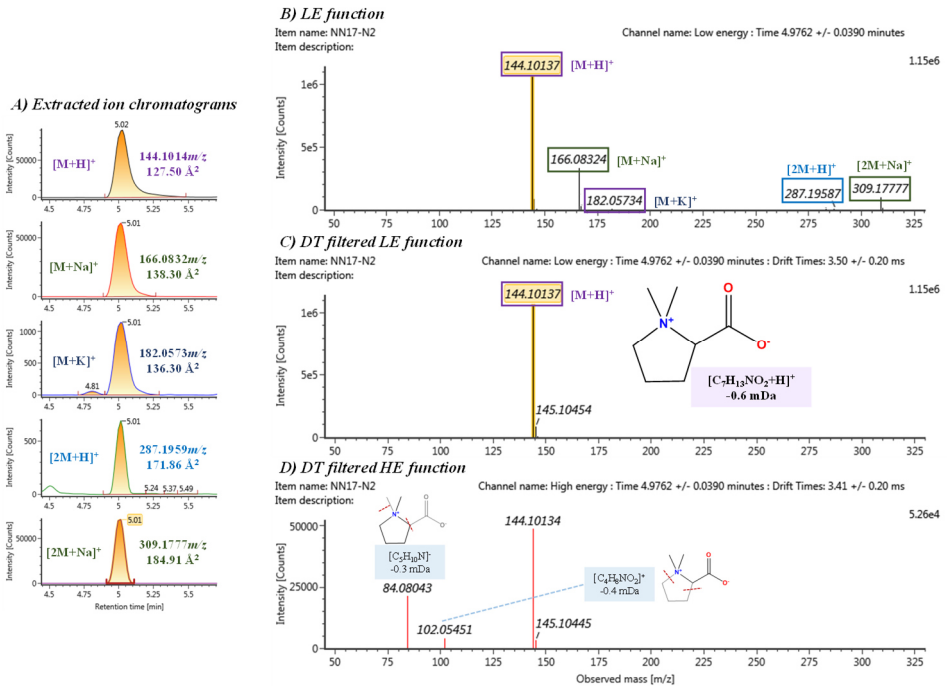


Figure 7. *Elucidation of marker 6 based on chromatograms and HDMS^E data obtained from HILIC pos analysis: (**A**) extracted ion chromatograms of all adduct and dimers found, along with the experimental m/z and CCS, (**B**) LE function, (**C**) DT filtered LE function and (**D**) DT filtered HE function. A proposed structure and fragmentation are shown.*

Indeed, stachydrine is one of the most commonly used dietary biomarkers for dietary assessment of citrus intake and it has even been related to the amount consumed [18,20]. Betonicine has also been detected in several studies, as well as other betonicines as Marker 4, tentatively elucidated as N-methyl-proline, due to the close retention time and fragmentation pattern. Marker 3 had fragmentation spectra similar to the other betaines reported. Some chemical structures were proposed based on the amino acid proline, although there are not enough data to propose a definitive structure (**Figure 4**). **Table 3** shows the available literature dealing with the biomarkers related to citrus intake.

Despite having linked all these compounds with orange consumption, it must be emphasised that they are usually exclusive to this type of food. Stachydrine is the most used biomarker as proof of citrus consumption. However, in some populations, it has also been detected as a marker for other foods such as tubers of the vegetable *Stachys affinis*, also known as Chinese artichoke [26]. Therefore, previous information of the foods more consumed in each population is important to select the most specific marker of orange intake, or even to use a combination of markers, and not just one, in selected populations. In order to achieve more confidence in tentative identification, a CCS predictor model machine-based in artificial neural network (ANN) developed for protonated ions by our group was employed [45]. The CCS values predicted for protonated ions are shown in **Table 2**. No prediction for the other adduct or dimers could be made as this tool is not yet developed for them. The relative errors obtained with this prediction tool were below 6 % for 95 % of all CCS values tested. Nevertheless, the results obtained for the phase II sulfate conjugated metabolites (of synephrine, hesperitine and N-methyltyramine) were between 5 % and 7 %. This higher error value is probably due to the lack of glucuronides and sulfate metabolites during the building of the CCS prediction model. Compared to other markers, good results were obtained, showing a prediction error below 1 % for stachidrine and betonicine and below 3 % in de case of N-Methyl-proline and Marker 3. The CCS predicted in the two structure proposals for marker 3 did not differ enough to decide between them. The CCS is used as an additional identification point that allows the identity confirmation. If the CCS is

different enough it will be able to distinguish between two isobaric and even isomeric compounds. Nevertheless, there is not enough information and CCS recompilation to apply this differentiation.

The increasing interest in the CCS as an additional molecular descriptor for untargeted metabolomic studies lies in its high capability to improve identification workflows. With IMS, post-ionization separation is undertaken based on the shape of the compounds and providing CCS as an additional physicochemical. In contrast to other parameters such as retention time, this value is highly reproducible since it is not influenced by the nature of the matrix or the separation employed. Therefore, it is of great importance to report these results, so that they can be included in compound libraries and, in this way, potentially improve the very time-consuming identification step in untargeted metabolomics studies.

Nevertheless, the ideal situation for unequivocally confirming the identity of the reported markers would be a comparison with a reference standard. Unfortunately, many of them are not commercially available, or are very expensive.

4. Conclusions

The potential of the untargeted metabolomics approach, along with the novelty of UHPLC-IMS-QTOF instruments for food intake biomarkers has been demonstrated with the tentative elucidation of 7 short-term markers of orange consumption (Clementines) in human plasma: Stachydrine, betonicine, N-methyl-proline, as well as phase II sulfate conjugate of synephrine, N-methyltyramine and hesperitin (only reported as biomarker of citrus intake in urine). Moreover, the recompilation in databases of CCS as an additional molecular descriptor, independent from the sample matrix nature and separation technique, will facilitate the future compound identification process. The markers have turned out to be indicators of citrus intake, not just orange. However, oranges (mainly *Citrus reticulata* and *Citrus sinensis*) are the fruits most consumed in the citrus group). The combined use of two or more of these markers can contribute to increase the specificity of the citrus intake.

Table 3. *Literature found on biomarkers related with the citrus intake.*

No.	Compound		Citrus fruit and/ or product consumed	Bio fluid sample	Analysis Platform	Literature on detection as metabolite	Literature demonstrating the presence in citrus
1	Synephrime hydrogen sulfate	Synephrine and/or other phase I/II metabolites	orange juice	urine	LC-MS/MS LC-ECD	[30,52]	[49,53]
2	N-methyltyramine hydrogen sulfate	N-methyltyramine and/or other phase I/II metabolites	Orange, grapefruit, orange juice	urine	LC-HRMS	[54]	[49,53]
3	Hesperitin hydrogen sulfate	Hesperitin and/or other phase I/II metabolites	Orange, grapefruit, orange juice	urine	LC-HRMS LC-MS/MS	[17,54]	[56]
4	N-methyl-proline		orange juice	urine	LC-HRMS	[21,25,27,28]	[55]
5	Betonicine		Orange, grapefruit, orange juice	plasma, urine	FIA-HRMS LC-HRMS LC-MS/MS	[17,19,25,28,54]	[55]
6	Stachydrine		Orange, grapefruit, orange juice, grapefruit juice	plasma, serum, urine	FIA-HRMS LC-M/MS LC-HRMS 1H NMR	[16,17,28,54,18–21,23–25,27]	[55]
7	unknown						

Author Contributions

Conceptualization, L.L.-B., T.P., J.V.S., F.J.L., O.C. and D.C.; methodology, L.L.-B., T.P., J.V.S. and D.C.; software, L.L.-B., J.V.S. and O.C.; validation, T.P., F.J.L., C.O.-A., E.M.A. and D.C.; formal analysis, L.L.-B., T.P. and D.C.; investigation, L.L.-B., C.O.-A. and E.M.A.; resources, T.P., F.J.L., J.V.S. and D.C.; data curation, L.L.-B., T.P., O.C. and D.C.; writing—original draft preparation, L.L.-B., T.P., O.C. and D.C.; writing—review and editing, L.L.-B., T.P., F.J.L., J.V.S., C.O.-A., E.M.A., O.C. and D.C.; supervision, T.P. and D.C.; funding acquisition, T.P., F.J.L., J.V.S., O.C. and D.C. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by Universitat Jaume I, predoctoral grant (L. Lacalle-Bergeron) number UJI 19I001/03 and grant number P1–1B2013–54; Ramon y Cajal Program (T. Portolés) of the Spanish Ministry of Economy and Competitiveness grant number RYC-2017-22525; the Spanish Ministry of Health (Instituto de Salud Carlos III) and the Ministerio de Economía y Competitividad-Fondo Europeo de Desarrollo Regional (FEDER) grants number CIBER 06/03, and SAF2016–80532-R; the Fundació La Marató de TV3 grant number 538/U/2016); and the Generalitat Valenciana grant number PROMETEO2017/017.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Picó, C.; Serra, F.; Rodríguez, A.M.; Keijer, J.; Palou, A. Biomarkers of nutrition and health: new tools for new approaches. Nutrients 2019, 11, 1092.
2. Naska, A.; Lagiou, A.; Lagiou, P. Dietary assessment methods in epidemiological research: Current state of the art and future prospects. F1000Research 2017, 6, 926.
3. Subar, A.F.; Freedman, L.S.; Tooze, J.A.; Kirkpatrick, S.I.; Boushey, C.; Neuhouser, M.L.; Thompson, F.E.; Potischman, N.; Guenther, P.M.; Tarasuk,

V.; et al. Addressing current criticism regarding the value of self-report dietary data. J. Nutr. 2015, 145, 2639–2645.
4. Falomir, Z.; Arregui, M.; Madueño, F.; Corella, D.; Coltell, Ó. Automation of food questionnaires in medical studies: a state-of-the-art review and future prospects. Comput. Biol. Med. 2012, 42, 964–974.
5. Conrad, J.; Nöthlings, U. Innovative approaches to estimate individual usual dietary intake in large-scale epidemiological studies. Proc. Nutr. Soc. 2017, 76, 213–219.
6. Sébédio, J.-L. Metabolomics, nutrition, and potential biomarkers of food quality, intake, and health status. In advances in food and nutrition research; Elsevier: Amsterdam, The Netherlands, 2017; Volume 82, pp. 83–116.
7. Brouwer-Brolsma, E.M.; Brennan, L.; Drevon, C.A.; van Kranen, H.; Manach, C.; Dragsted, L.O.; Roche, H.M.; Andres-Lacueva, C.; Bakker, S.J.L.; Bouwman, J.; et al. Combining traditional dietary assessment methods with novel metabolomics techniques: Present efforts by the Food Biomarker Alliance. Proc. Nutr. Soc. 2017, 76, 619–627.
8. O'Gorman, A.; Brennan, L. The role of metabolomics in determination of new dietary biomarkers. Proc. Nutr. Soc. 2017, 76, 295–302.
9. Woodside, J.V.; Draper, J.; Lloyd, A.; McKinley, M.C. Use of biomarkers to assess fruit and vegetable intake. Proc. Nutr. Soc. 2017, 76, 308–315.
10. Dragsted, L.O.; Gao, Q.; Scalbert, A.; Vergères, G.; Kolehmainen, M.; Manach, C.; Brennan, L.; Afman, L.A.; Wishart, D.S.; Andres Lacueva, C.; et al. Validation of biomarkers of food intake—critical assessment of candidate biomarkers. Genes Nutr. 2018, 13, 14.
11. Ulaszewska, M.M.; Weinert, C.H.; Trimigno, A.; Portmann, R.; Andres Lacueva, C.; Badertscher, R.; Brennan, L.; Brunius, C.; Bub, A.; Capozzi, F.; et al. Nutrimetabolomics: An integrative action for metabolomic analyses in human nutritional studies. Mol. Nutr. Food Res. 2019, 63, 1–38.
12. Praticò, G.; Gao, Q.; Scalbert, A.; Vergères, G.; Kolehmainen, M.; Manach, C.; Brennan, L.; Pedapati, S.H.; Afman, L.A.; Wishart, D.S.; et al. Guidelines for Biomarker of Food Intake Reviews (BFIRev): How to conduct an extensive literature search for biomarker of food intake discovery. Genes Nutr. 2018, 13, 1–14.
13. Maruvada, P.; Lampe, J.W.; Wishart, D.S.; Barupal, D.; Chester, D.N.; Dodd, D.; Djoumbou-Feunang, Y.; Dorrestein, P.C.; Dragsted, L.O.; Draper, J.; et al.

Perspective: Dietary biomarkers of intake and exposure—exploration with omics approaches. Adv. Nutr. 2019, 200–215.
14. Harris, J.E.; Raynor, H.A. Crossover designs in nutrition and dietetics research. J. Acad. Nutr. Diet. 2017, 117, 1023–1030.
15. Guasch-Ferré, M.; Bhupathiraju, S.N.; Hu, F.B. Use of Metabolomics in improving assessment of dietary intake. Clin. Chem. 2018, 64, 82–98.
16. Heinzmann, S.S.; Brown, I.J.; Chan, Q.; Bictash, M.; Dumas, M.E.; Kochhar, S.; Stamler, J.; Holmes, E.; Elliott, P.; Nicholson, J.K. Metabolic profiling strategy for discovery of nutritional biomarkers: Proline betaine as a marker of citrus consumption. Am. J. Clin. Nutr. 2010, 92, 436–443.
17. Lloyd, A.J.; Beckmann, M.; Favé, G.; Mathers, J.C.; Draper, J. Proline betaine and its biotransformation products in fasting urine samples are potential biomarkers of habitual citrus fruit consumption. Br. J. Nutr. 2011, 106, 812–824.
18. Gibbons, H.; Michielsen, C.J.R.; Rundle, M.; Frost, G.; McNulty, B.A.; Nugent, A.P.; Walton, J.; Flynn, A.; Gibney, M.J.; Brennan, L. Demonstration of the utility of biomarkers for dietary intake assessment; proline betaine as an example. Mol. Nutr. Food Res. 2017, 61, 1–8.
19. Atkinson, W.; Downer, P.; Lever, M.; Chambers, S.T.; George, P.M. Effects of orange juice and proline betaine on glycine betaine and homocysteine in healthy male subjects. Eur. J. Nutr. 2007, 46, 446–452.
20. Lang, R.; Lang, T.; Bader, M.; Beusch, A.; Schlagbauer, V.; Hofmann, T. High-Throughput Quantitation of Proline Betaine in Foods and Suitability as a Valid Biomarker for Citrus Consumption. J. Agric. Food Chem. 2017, 65, 1613–1619.
21. Guertin, K.A.; Moore, S.C.; Sampson, J.N.; Huang, W.Y.; Xiao, Q.; Stolzenberg-Solomon, R.Z.; Sinha, R.; Cross, A.J. Metabolomics in nutritional epidemiology: Identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. Am. J. Clin. Nutr. 2014, 100, 208–217.
22. D'Angelo, S.; Gormley, I.C.; McNulty, B.A.; Nugent, A.P.; Walton, J.; Flynn, A.; Brennan, L. Combining biomarker and food intake data: Calibration equations for citrus intake. Am. J. Clin. Nutr. 2019, 110, 977–983.
23. May, D.H.; Navarro, S.L.; Ruczinski, I.; Hogan, J.; Ogata, Y.; Schwarz, Y.; Levy, L.; Holzman, T.; McIntosh, M.W.; Lampe, J.W. Metabolomic profiling

of urine: Response to a randomised, controlled feeding study of select fruits and vegetables, and application to an observational study. Br. J. Nutr. 2013, 110, 1760–1770.

24. Mazzilli, K.M.; McClain, K.M.; Lipworth, L.; Playdon, M.C.; Sampson, J.N.; Clish, C.B.; Gerszten, R.E.; Freedman, N.D.; Moore, S.C. Identification of 102 Correlations between Serum Metabolites and Habitual Diet in a Metabolomics Study of the Prostate, Lung, Colorectal, and Ovarian Cancer Trial. J. Nutr. 2019, 1–10.
25. Rangel-Huerta, O.D.; Aguilera, C.M.; Perez-de-la-Cruz, A.; Vallejo, F.; Tomas-Barberan, F.; Gil, A.; Mesa, M.D. A serum metabolomics-driven approach predicts orange juice consumption and its impact on oxidative stress and inflammation in subjects from the BIONAOS study. Mol. Nutr. Food Res. 2017, 61, 1–12.
26. Garcia-Perez, I.; Posma, J.M.; Gibson, R.; Chambers, E.S.; Hansen, T.H.; Vestergaard, H.; Hansen, T.; Beckmann, M.; Pedersen, O.; Elliott, P.; et al. Objective assessment of dietary patterns by use of metabolic phenotyping: A randomised, controlled, crossover trial. Lancet Diabetes Endocrinol. 2017, 5, 184–195.
27. Wang, Y.; Gapstur, S.M.; Carter, B.D.; Hartman, T.J.; Stevens, V.L.; Gaudet, M.M.; McCullough, M.L. Untargeted metabolomics identifies novel potential biomarkers of habitual food intake in a cross-sectional study of postmenopausal women. J. Nutr. 2018, 148, 932–943.
28. Playdon, M.C.; Sampson, J.N.; Cross, A.J.; Sinha, R.; Guertin, K.A.; Moy, K.A.; Rothman, N.; Irwin, M.L.; Mayne, S.T.; Stolzenberg-Solomon, R.; et al. Comparing metabolite profiles of habitual diet in serum and urine. Am. J. Clin. Nutr. 2016, 104, 776–789.
29. Milenkovic, D.; Morand, C.; Cassidy, A.; Konic-Ristic, A.; Tomás-Barberán, F.; Ordovas, J.M.; Kroon, P.; De Caterina, R.; Rodriguez-Mateos, A. Interindividual Variability in Biomarkers of Cardiometabolic Health after Consumption of Major Plant-Food Bioactive Compounds and the Determinants Involved. Adv. Nutr. 2017, 8, 558–570.
30. Bader, M.; Lang, T.; Lang, R.; Hofmann, T. Synephrine as a Specific Marker for Orange Consumption. J. Agric. Food Chem. 2017, 65, 4853–4858.

31. Collins, C.; McNamara, A.E.; Brennan, L. Role of metabolomics in identification of biomarkers related to food intake. Proc. Nutr. Soc. 2019, 78, 189–196.
32. Puiggròs, F.; Solà, R.; Bladé, C.; Salvadó, M.J.; Arola, L. Nutritional biomarkers and foodomic methodologies for qualitative and quantitative analysis of bioactive ingredients in dietary intervention studies. J. Chromatogr. A 2011, 1218, 7399–7414.
33. Dettmer, K.; Aronov, P.A.; Hammock, B.D. Mass Spectrometry-Based Metabolomics. Mass Spectrom. Rev. 2007, 26, 51–78.
34. Cevallos-Cevallos, J.M.; Reyes-De-Corcuera, J.I.; Etxeberria, E.; Danyluk, M.D.; Rodrick, G.E. Metabolomic analysis in food science: A review. Trends Food Sci. Technol. 2009, 20, 557–566.
35. Nicholson, J.K.; Wilson, I.D. High resolution proton magnetic resonance spectroscopy of biological fluids. Prog. Nucl. Magn. Reson. Spectrosc. 1989, 21, 449–501. [Google Scholar] [CrossRef]
36. Nicholson, J.K.; Lindon, J.C. Metabonomics. Nature 2008, 455, 1054–1056.
37. Estabrooks, P.A.; Serrano, E.; Savla, J.; Hedrick, V.E.; Davy, B.M.; Dietrich, A.M. Dietary biomarkers: Advances, limitations and future directions. Nutr. J. 2012, 11, 1. [Google Scholar]
38. De Vijlder, T.; Valkenborg, D.; Lemière, F.; Romijn, E.P.; Laukens, K.; Cuyckens, F. A tutorial in small molecule identification via electrospray ionization-mass spectrometry: The practical art of structural elucidation. Mass Spectrom. Rev. 2017, 37, 607–629.
39. Ardrey, R.E. Liquid Chromatography-Mass Spectrometry: An Introduction; Willey Editorial: Hoboken, NJ, USA, 2003.
40. Fenaille, F.; Barbier Saint-Hilaire, P.; Rousseau, K.; Junot, C. Data acquisition workflows in liquid chromatography coupled to high resolution mass spectrometry-based metabolomics: Where do we stand? J. Chromatogr. A 2017, 1526, 1–12.
41. Kaufmann, A.; Walker, S. Comparison of linear intrascan and interscan dynamic ranges of Orbitrap and ion-mobility time-of-flight mass spectrometers. Rapid Commun. Mass Spectrom. 2017, 31, 1915–1926.
42. Gabelica, V.; Marklund, E. Fundamentals of ion mobility spectrometry. Curr. Opin. Chem. Biol. 2018, 42, 51–59.

43. D'Atri, V.; Causon, T.; Hernandez-Alba, O.; Mutabazi, A.; Veuthey, J.L.; Cianferani, S.; Guillarme, D. Adding a new separation dimension to MS and LC–MS: What is the utility of ion mobility spectrometry? J. Sep. Sci. 2018, 41, 20–67.
44. Mairinger, T.; Causon, T.J.; Hann, S. The potential of ion mobility–mass spectrometry for non-targeted metabolomics. Curr. Opin. Chem. Biol. 2018, 42, 9–15.
45. Bijlsma, L.; Bade, R.; Celma, A.; Mullin, L.; Cleland, G.; Stead, S.; Hernandez, F.; Sancho, J.V. Prediction of collision cross-section values for small molecules: application to pesticide residue analysis. Anal. Chem. 2017, 89, 6583–6589.
46. Barragán, R.; Coltell, O.; Portolés, O.; Asensio, E.M.; Sorlí, J.V.; Ortega-Azorín, C.; González, J.I.; Sáiz, C.; Fernández-Carrión, R.; Ordovas, J.M.; et al. Bitter, sweet, salty, sour and umami taste perception decreases with age: Sex-specific analysis, modulation by genetic variants and taste-preference associations in 18 to 80 year-old subjects. Nutrients 2018, 10, 1539.
47. Rossato, L.G.; Costa, V.M.; Limberger, R.P.; de Bastos, M.L.; Remião, F. Synephrine: From trace concentrations to massive consumption in weight-loss. Food Chem. Toxicol. 2011, 49, 8–16.
48. Parhiz, H.; Roohbakhsh, A.; Soltani, F.; Rezaee, R.; Iranshahi, M. Antioxidant and anti-inflammatory properties of the citrus flavonoids hesperidin and hesperetin: An updated review of their molecular mechanisms and experimental models. Phyther. Res. 2015, 29, 323–331.
49. Haaz, S.; Fontaine, K.R.; Cutter, G.; Limdi, N.; Perumean-Chaney, S.; Allison, D.B. Citrus aurantium and synephrine alkaloids in the treatment of overweight and obesity: An update. Obes. Rev. 2006, 7, 79–88.
50. Matsumoto, H.; Ikoma, Y.; Sugiura, M.; Yano, M.; Hasegawa, Y. Identification and quantification of the conjugated metabolites derived from orally administered hesperidin in rat plasma. J. Agric. Food Chem. 2004, 52, 6653–6659.
51. Yang, C.Y.; Tsai, S.Y.; Chao, P.D.L.; Yen, H.F.; Chien, T.M.; Hsiu, S.L. Determination of hesperetin and its conjugate metabolites in serum and urine. J. Food Drug Anal. 2002, 10, 143–148.
52. Kusu, F.; Matsumoto, K.; Arai, K.; Takamura, K. Determination of synephrine enantiomers in food and conjugated synephrine in urine by high-

performance liquid chromatography with electrochemical detection. Anal. Biochem. 1996, 235, 191–194.

53. Nelson, B.C.; Putzbach, K.; Sharpless, K.E.; Sander, L.C. Mass spectrometric determination of the predominant adrenergic protoalkaloids in bitter orange (*Citrus aurantium*). J. Agric. Food Chem. 2007, 55, 9769–9775.
54. Pujos-Guillot, E.; Hubert, J.; Martin, J.F.; Lyan, B.; Quintana, M.; Claude, S.; Chabanas, B.; Rothwell, J.A.; Bennetau-Pelissero, C.; Scalbert, A.; et al. Mass spectrometry-based metabolomics for the discovery of biomarkers of fruit and vegetable intake: citrus fruit as a case study. J. Proteome Res. 2013, 12, 1645–1659.
55. Gattuso, G.; Barreca, D.; Gargiulli, C.; Leuzzi, U.; Caristi, C. Flavonoid composition of citrus juices. Molecules 2007, 12, 1641–1673.
56. Servillo, L.; Giovane, A.; Balestrieri, M.L.; Bata-Csere, A.; Cautela, D.; Castaldo, D. Betaines in fruits of citrus genus plants. J. Agric. Food Chem. 2011, 59, 9410–9416.

Supplementary Information

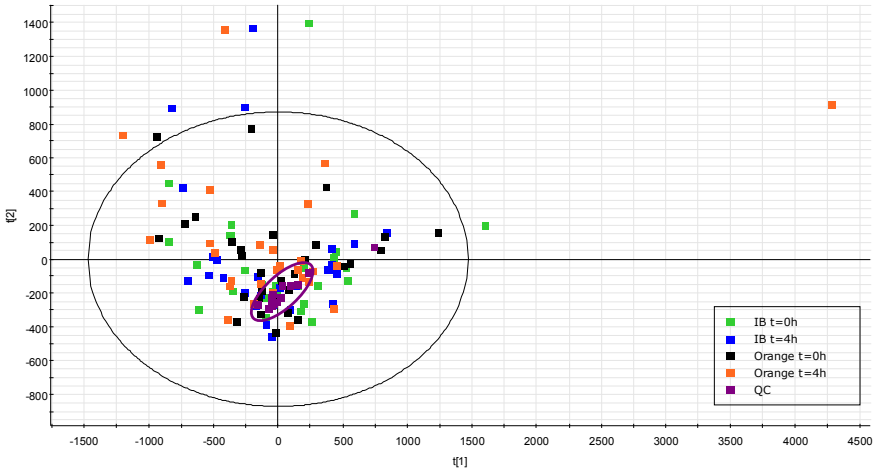


Figure S1. *PCA score plot component 1 vs component 2 of HILIC in negative ionization mode, explaining the 41.8 % and 14.7 % of the variance, respectively. The purple points (14 QC samples) are grouped and centred in the plot. 4 plasma samples were obtained from each of the 30 participants, obtaining 30 samples per group. 120 plasma samples were analysed in total. The samples obtained at t = 0 h and t = 4 h after the intake of an isocaloric beverage (IB) and the samples obtained at t = 0 h and t = 4 h after orange intake are coloured in green, blue, black and orange, respectively.*

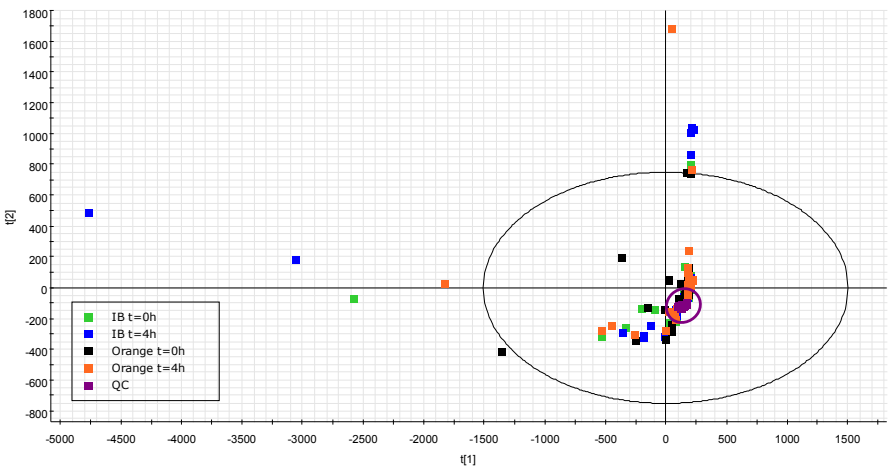


Figure S2. *PCA score plot component 1 vs component 2 of RP in positive ionization mode, explaining the 47.1 % and 11.7 % of the variance, respectively. The purple points (14 QC samples) are grouped and centred in the plot. 4 plasma samples were obtained from each of the 30 participants, obtaining 30 samples per group. 120 plasma samples were analysed in total. The samples obtained at t = 0 h and t = 4 h after the intake of an isocaloric beverage (IB) and the samples obtained at t = 0 h and t = 4 h after orange intake are coloured in green, blue, black and orange, respectively.*

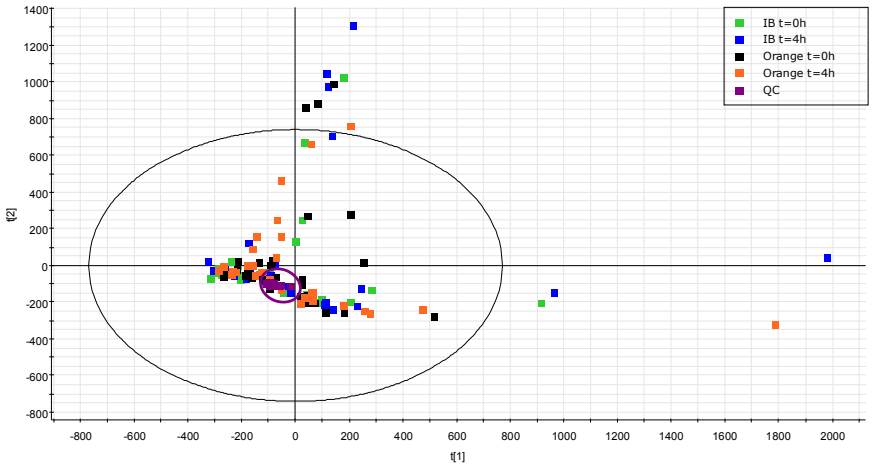


Figure S3. *PCA score plot component 1 vs component 2 of RP in negative ionization mode, explaining the 28.5 % and 26.7 % of the variance, respectively. The purple points (14 QC samples) are grouped and centred in the plot. 4 plasma samples were obtained from each of the 30 participants, obtaining 30 samples per group. 120 plasma samples were analysed in total. The samples obtained at t = 0 h and t = 4 h after the intake of an isocaloric beverage (IB) and the samples obtained at t = 0 h and t = 4 h after orange intake are coloured in green, blue, black and orange, respectively.*

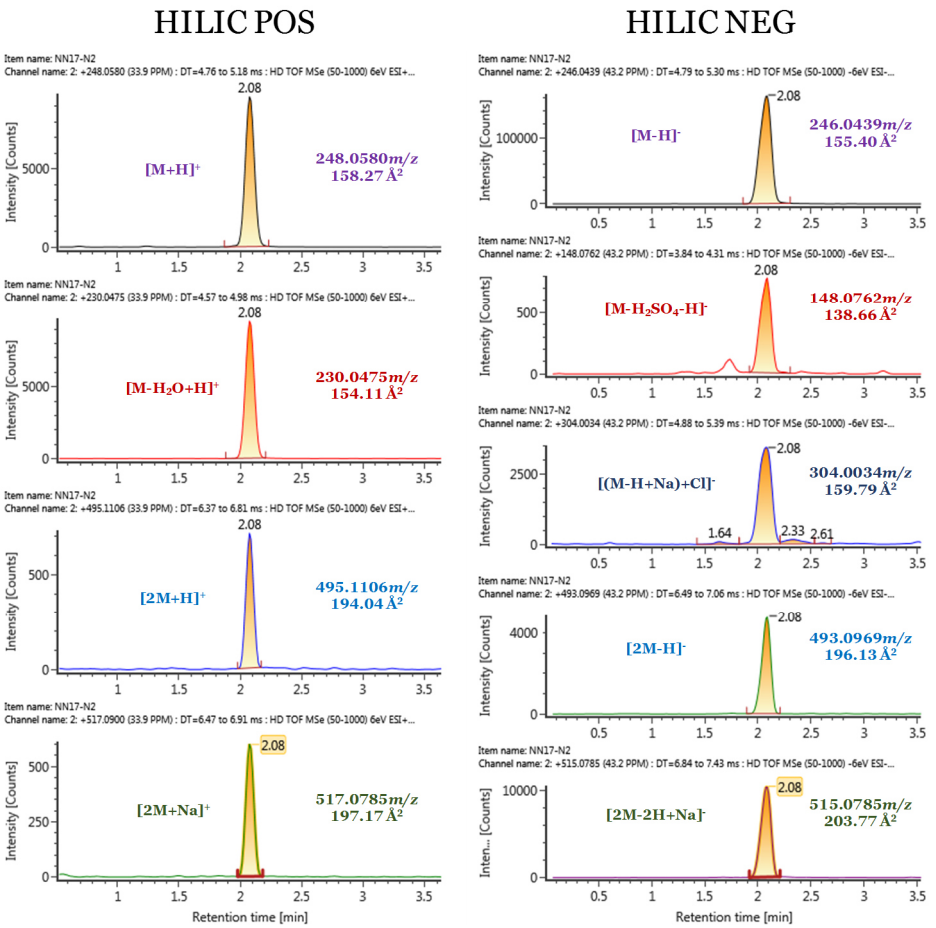


Figure S6. *Extracted ion chromatograms of ions present in the LE function of HILIC pos and HILIC neg analysis. The experimental RT and CCS for each ion form are shown.*

3.3 Medium-term biomarkers of orange intake

Dietary biomarkers are desirable for their ability to assess more accurately the nutritional intake *versus* self-reported methodologies, although its usefulness will depend on the period of time of monitoring.

BFI can be classified as short-term to reflect the intake over the past hours/days, medium-term over the weeks/months; and long-term markers to reflect the intake over the months and even years (Hedrick et al., 2012). This will depend on if it is an exogenous or metabolised compound, its half-life in the body and if it can be bioaccumulated, among other characteristics. The selection of the sample could also be important factor for the BFI monitorization. For example, a hydrophilic exogenous compound that does not undergoes any biotransformation or degradation in the body has high probability to be rapidly excreted, being a good short-term BFI in urine. On the other hand, BFIs accumulated in blood, hair or nails are good as long-term or repeated exposure markers (Dragsted et al., 2018).

Due to the less complex study design, ease of obtaining adequate samples and validation, the majority of the studies have been mainly focused on short-term exposure.

Along with the previous study compiled in **Scientific Article III**, a preliminary study of medium-term biomarkers of orange intake was performed. Contrary to the previous one, it was not a cross-over trial and therefore the great advantage of reducing inter-subject variations was lost by not passing the same subject through all study stages. In fact, only 16 subjects out of 30 who participated in the previous study agreed to participate in this second phase (aged 24.8 ± 3.3 years, 6 females and 10 males). Of those 16, 8 participants agreed to consume approximately 1 orange per day during a month ("orange" group), while the other 8 subjects assent to avoid citrus during the same period of time ("control" group). At the end of the month, a plasma sample was taken.

When a crossover study cannot be contemplated, as in the case of observational studies, a higher number of samples is normally required in order to reduce the inter-subject variations. For this reason, this untargeted metabolomic study for the discovery of medium-term biomarkers of orange

consumption was considered exploratory and preliminary due to the small number of subjects participating.

Regarding to the sample treatment, instrumentation and analysis conditions applied to the samples from the medium-term study, they were processed in the same way and at the same time as the samples from short-term study. As data processing was also performed along with short-term samples, and prior to any kind of sample grouping, the same number of features as reported in the previous study were detected: 6951, 6238, 5283 and 4479 ions in RP pos, RP neg, HILIC pos and HILIC neg, respectively. Statistical analysis and elucidation workflow were also identical to the described in the **Scientific Article III**.

The first statistical analysis to be performed was the principal component analysis (PCA). The control of the correct acquisition of the sample based on the correct centring of the QCs (quality control samples, pooled sample from the all set of samples analysed) in the PCA can be observed in **Figure 1** of **Scientific Article III**. The data sets were then reduced to 2325, 1480, 1314 and 676 in RP pos, RP neg, HILIC pos and HILIC neg respectively, after the removal of compounds that exhibit a poor stability with a relative standard deviation (% RSD) below 30 % within the QCs.

Then, focusing on the medium-term study, “control” and “orange” samples were grouped, and one-way ANOVA calculation was applied assuming the independence of each sample followed by a false discovery rate (FDR), being only 51, 11, 58 and 3 statistically significant features for RP pos, RP neg, HILIC pos and HILIC neg, respectively. Due to the scarce number of features that accomplish the ANOVA threshold, only RP pos and HILIC pos data was used to perform further statistical analysis.

Then a second PCA, only focusing in the medium-term study, was performed. The PCA score plots of the first two components are shown in **Figure 1**, where it can be observed a differentiation between the two groups in both data sets. Outliers were not observed attending to the 95 % of confidence limit of the Hotelling’s T2 Range.

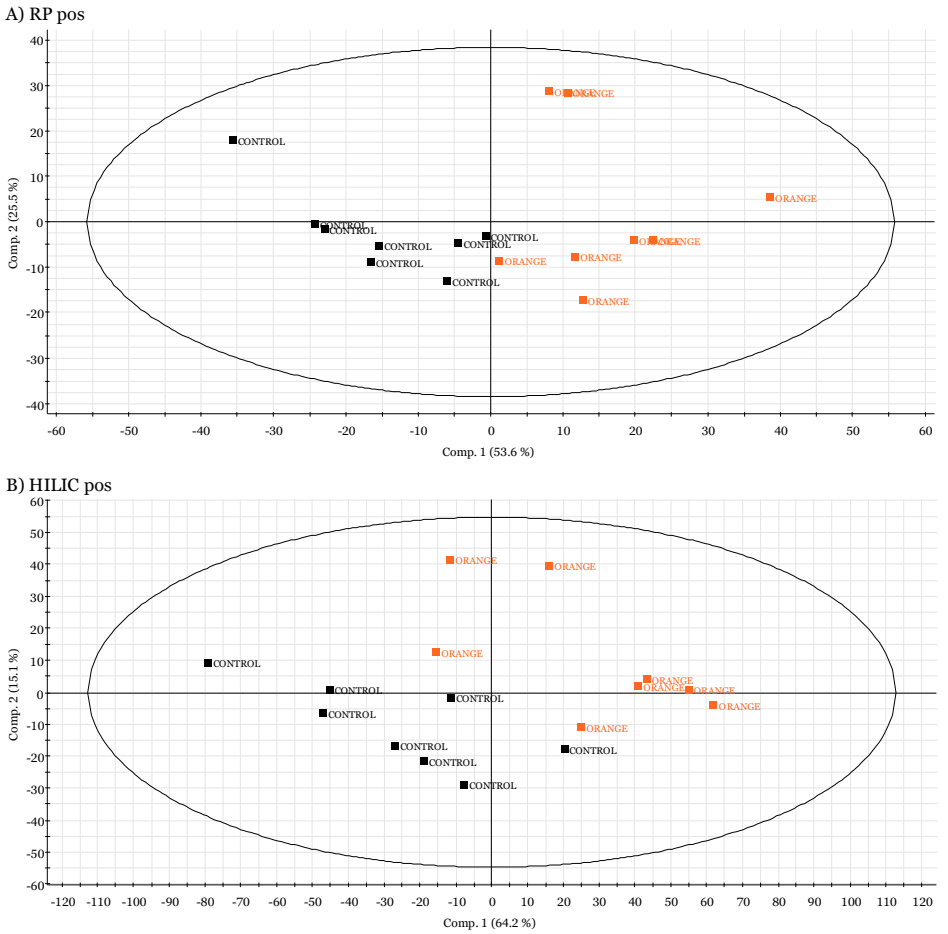


Figure 1. *PCA score plot component 1 vs component 2 for A) RP and B) HILIC, both in positive ionization mode. The two groups of 8 samples each were analysed, coming from: control participates with no consumption of citrus during one moth (CONTROL ■) and participants that have consumed one orange daily (ORANGE ■). The explained was both 79 % for RP pos and HILIC pos, respectively.*

In order to highlight the most significant features, the supervised multivariate statistical orthogonal partial least square-discriminant analysis (OPLS-DA) was employed, where the "control" group was faced to the "orange" group, obtaining: R^2 = 82 % and Q^2 = 64 % for RP pos and R^2 = 82 % and Q^2 = 73 % for HILIC pos. From the OPLS-DA the S-plot was generated in order to visualise and select the most relevant features. A cut-off of *p(corr)* ≥ |0.6| and *p[1] loadings* ≥ |0.1| were employed to this end in each data set,

and 1 and 5 features for RP pos and HILIC pos, respectively, were highlighted as markers for the "orange" group.

Following the same process as explained in the short-term study article, the highlighted features were carefully reviewed, and it was intended to give a tentative identification.

When reviewing the possible markers, 3 of the HILIC pos features were eliminated as putative markers because of the bad peak shape and poor mass spectra in the raw data. It is not unusual that the automatization of peak picking and retention time alignment returns some "false peaks" in the data set created. Usually those "peaks" are part of the noise or coming from a not well-defined peak, and they could be reduced by changing some parameters during the peak picking process. This verification is too complicated to be performed before the statistical analysis to the whole data set. For this reason, it is compulsory to validate a compound as a marker and provide an identity.

In this regard, a prominent question is "why such features are highlighted in the statistical analysis?". As it can be seen above, not many features passed the ANOVA threshold, and the following statistical analysis might suffer of over fitting an give more importance to these kinds of features. Regarding other statistical parameters, the eliminated features presented a fold change (i.e. the largest deviation of the trend line from the base level) below 1.2, it is usually recommended a fold change ≥ 1.5, showing that, although there is a slight difference , it is not really relevant. Moreover, in the short-term study all the elucidated markers presented a fold change > 10, some of them even higher than 100, demonstrating a big difference between the faced groups.

Turning to the 3 features left, the feature HILICpos_3.44_735.6012n presented a *p(corr)* of 0.78 and fold change of 1.51. Two adducts were associated to this compound ($[M+H]^+$ and $[M+Na]^+$) and thanks to the availability of spectral libraries, this compound was tentatively identified as betaine lipid diacylglyceryltrimethylhomoserine (16:0/ 18:2) (DGTS (16:0/18:2)), with an elemental composition $C_{44}H_{81}NO_7$ (-0.1 mDa error) and CCS of 295.32 $Å^2$. The high energy spectra showed a typical product ion of DGTS lipids at *m/z* 236.1494, which corresponded to the ion $C_{10}H_{22}O_5N^+$

(-0.4 mDa) generated from the loss of both acyl chains (Alves et al., 2019; Li et al., 2017). Betaine lipids are common in lower plants, algae and in some non-photosynthetic microorganisms as bacteria or fungus, being important constituents of their cell membranes (Sato, 1992). As far as we know, this type of lipids are not generated by human metabolism but can be generated by some bacteria. Therefore, it is necessary to stablish the origin of this metabolite: if it is part of the ingested metabolites, even not being one of the markers from the short-term study, or on the contrary, if it could come from the metabolism of the microbiota inherent in the human body. This type of polar lipids highly present in microalgae have been studied for their potential anti-inflammatory properties, and therefore, the algae extract as possible functional foods (Conde et al., 2021; da Costa et al., 2021).

Finally, the most differentiating feature was found in both HILIC pos (HILICpos_5.02_143.0942n with *p(corr)* 0.97, fold change 4.53) and RP pos (RPpos_0.64_143.0947n with *p(corr)* 0.93, fold change 5.46), being the same compound. It corresponded to the compound already highlighted in the short-term study and tentatively identified as Stachydrine (Proline Betaine), the biomarker most used for the assessment of citrus intake as explained in the previous study (**Scientific Article III**).

Figure 2 shows the differences in abundance obtained between the short-term and medium-term study of orange intake. While "short-term control" (referenced as "Orange t=0 h" in **Scientific Article III**), "medium-term control" and "medium-term orange" groups of samples were collected in the morning after 8 h fasting, "short-term orange" (referend as "Orange t=4 h" in **Scientific Article III)** groups of samples were obtained 4 h after the consumption of 500 g peeled oranges, and no other food was allowed during this time. As it can be observed, the abundance of stachydrine obtained in the short-term study was four times higher than the one obtained for the medium-term study. This is probably due to, as it has been demonstrated in other studies, stachydrine has a relatively short half-life (Gibbons et al., 2017). Therefore, it seems logical that after a minimum of 8 h without having consumed any citrus, the level of this compound decreases significantly. Nevertheless, it is still a good marker of citrus consumption at plasmatic level during this period of time. This compound is naturally found

at lower level in the human body (Lever et al., 1994). In fact, dimers ($[2M+H]^+$ and $[2M+Na]^+$), and even the adduct $[M+K]^+$, are usually formed when the concentration of this compound is relatively important. However, in the control groups of both studies only the $[M+H]^+$ and $[M+Na]^+$ were detected, which denotes the low concentration found.

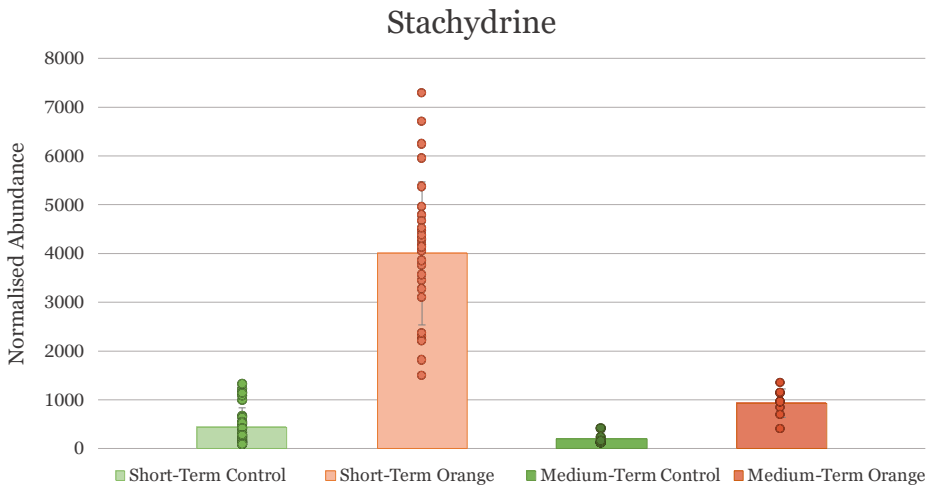


Figure 2. *Comparison of the normalised abundance of Stachydrine in both studies. In green shades the average abundance obtained from the control samples in short-term and medium-term study; and in orange shades the average abundance for the samples after the orange intake of short-term and medium-term study.*

In this study it has been demonstrated the breakthrough of IMS implementation in HRMS instruments in the medium-term biomarkers identification process. On the other hand, despite the fact that the work carried out is very useful at exploratory level, the results obtained showed that additional knowledge and a more elaborated experiment design are required to obtain robust medium/long- term BFI of orange. The main changes advisable to apply would be to perform a cross-over trial as in the short-term study; and increase the range of participants and lifestyle. In addition, the ideal situation will be to increase the sampling points during the month, to obtain information about the evolution of the markers; and at least 1 sample/month during a year to assess their validity as long-term BFI.

3.4 References

Alves, E., Melo, T., Barros, M. P., Domingues, M. R. M., & Domingues, P. (2019). Lipidomic profiling of the olive (olea europaea L.) fruit towards its valorisation as a functional Food: in-depth identification of triacylglycerols and polar lipids in portuguese olives. Molecules, 24(14), 2555. https://doi.org/10.3390/molecules24142555

Bijlsma, L., Bade, R., Celma, A., Mullin, L., Cleland, G., Stead, S., Hernandez, F., & Sancho, J. V. (2017). Prediction of Collision Cross-Section Values for Small Molecules: Application to Pesticide Residue Analysis. Analytical Chemistry, 89(12), 6583–6589. https://doi.org/10.1021/acs.analchem.7b00741

Conde, T. A., Zabetakis, I., Tsoupras, A., Medina, I., Costa, M., Silva, J., Neves, B., Domingues, P., & Domingues, M. R. (2021). Microalgal lipid extracts have potential to modulate the inflammatory response: A critical review. International Journal of Molecular Sciences, 22(18), 9825. https://doi.org/10.3390/ijms22189825

da Costa, E., Melo, T., Reis, M., Domingues, P., Calado, R., Abreu, M. H., & Domingues, M. R. (2021). Polar lipids composition, antioxidant and anti-inflammatory activities of the atlantic red seaweed grateloupia turuturu. Marine Drugs, 19(8), 414. https://doi.org/10.3390/md19080414

Dragsted, L. O., Gao, Q., Scalbert, A., Vergères, G., Kolehmainen, M., Manach, C., Brennan, L., Afman, L. A., Wishart, D. S., Andres Lacueva, C., Garcia-Aloy, M., Verhagen, H., Feskens, E. J. M., & Praticò, G. (2018). Validation of biomarkers of food intake—critical assessment of candidate biomarkers. Genes & Nutrition, 13(1), 14. https://doi.org/10.1186/s12263-018-0603-9

Gibbons, H., Michielsen, C. J. R., Rundle, M., Frost, G., McNulty, B. A., Nugent, A. P., Walton, J., Flynn, A., Gibney, M. J., & Brennan, L. (2017). Demonstration of the utility of biomarkers for dietary intake assessment; proline betaine as an example. Molecular Nutrition & Food Research, 61(10), 1700037. https://doi.org/10.1002/mnfr.201700037

Gomez-Delgado, F., Katsiki, N., Lopez-Miranda, J., & Perez-Martinez, P. (2021). Dietary habits, lipoprotein metabolism and cardiovascular disease: From individual foods to dietary patterns. Critical Reviews in Food Science and Nutrition, 61(10), 1651–1669. https://doi.org/10.1080/10408398.2020.1764487

Hedrick, V. E., Dietrich, A. M., Estabrooks, P. A., Savla, J., Serrano, E., & Davy, B. M. (2012). Dietary biomarkers: advances, limitations and future directions. Nutrition Journal, 11(1), 109. https://doi.org/10.1186/1475-2891-11-109

Kim, H., & Rebholz, C. M. (2021). Metabolomic biomarkers of healthy dietary patterns and cardiovascular outcomes. Current Atherosclerosis Reports, 23(6), 26. https://doi.org/10.1007/s11883-021-00921-8

Lever, M., Sizeland, P. C. B., Bason, L. M., Hayman, C. M., & Chambers, S. T. (1994). Glycine betaine and proline betaine in human blood and urine. Biochimica et Biophysica Acta (BBA) - General Subjects, 1200(3), 259–264. https://doi.org/10.1016/0304-4165(94)90165-1

Li, Y., Lou, Y., Mu, T., Xu, J., Zhou, C., & Yan, X. (2017). Simultaneous structural identification of diacylglyceryl-N-trimethylhomoserine (DGTS) and diacylglycerylhydroxymethyl-N,N,N-trimethyl-β-alanine (DGTA) in microalgae using dual Li+/H+ adduct ion mode by ultra-performance liquid chromatography/quadrupole ti. Rapid Communications in Mass Spectrometry, 31(5), 457–468. https://doi.org/10.1002/rcm.7818

O'Gorman, A., & Brennan, L. (2017). The role of metabolomics in determination of new dietary biomarkers. Proceedings of the Nutrition Society, 76(3), 295–302. https://doi.org/10.1017/S0029665116002974

Sato, N. (1992). Betaine lipids. The Botanical Magazine Tokyo, 105(1), 185–197. https://doi.org/10.1007/BF02489414

Vázquez-Manjarrez, N., Ulaszewska, M., Garcia-Aloy, M., Mattivi, F., Praticò, G., Dragsted, L. O., & Manach, C. (2020). Biomarkers of intake for tropical fruits. Genes & Nutrition, 15(1), 11. https://doi.org/10.1186/s12263-020-00670-4

Wu, B., Wei, F., Xu, S., Xie, Y., Lv, X., Chen, H., & Huang, F. (2021). Mass spectrometry-based lipidomics as a powerful platform in foodomics research. Trends in Food Science and Technology, 107(November 2020), 358–376. https://doi.org/10.1016/j.tifs.2020.10.045

Chapter 4.
Food and health effects

Chapter 4. Food and health effects

4.1. Introduction

4.2. Scientific article IV

"The potential of ion mobility separation in combination with high resolution mass spectrometry for the identification of biomarkers highlighted by untargeted metabolomics: the effects of pterostilbene and resveratrol consumption in liver steatosis, animal model"

Food Chemistry (2021) Submitted

4.3 Discussion of the results

4.4 References

4.1 Introduction

In last decades, nutrition research has shifted from a focus on providing enough food to nourish people to being a potential health promoter. Scientific research has not only been limited to food-related areas such as quality management, safety and traceability, but intrinsically related to health (Braconi et al., 2018). Nowadays, one of the main objectives in the nutrition field is to increase the knowledge about the role of ingested compounds at the molecular level (i.e. interactions with genes, proteins and/or metabolome). Thanks to this knowledge, advanced strategies could be developed to manipulate cellular functions through diet or the ingestion of certain foods could be advised, with its consequent impact on health (García-Cañas et al., 2010). In addition, it has increased the interest of the food industry in the development of new food products containing high concentrations of compounds with reported biological activity beneficial to health (e.g. nutraceuticals, food supplements or functional foods) (Fernández-Ochoa et al., 2021).

On one hand, the role of metabolomics in this area is focused on the food products characterization with special interest in putative bioactive components (Mancano et al., 2018). This has been carried out mainly by targeted metabolomics based on the investigation of families of compounds with reported bioactivity, such as polyphenols (Puiggròs et al., 2011). Although the number of untargeted metabolomic studies for this purpose is gradually increasing, it should always be accompanied by consecutive clinical trials to determine the bioactivity of the reported markers (Claus, 2014). On the other hand, untargeted metabolomics has been widely applied in the research of metabolic alterations produced by diet, the intake of certain nutrients or bioactive food constituents. Biomarkers derived from these studies could be both exposure biomarkers (exogenous or metabolised, employed for the assessment of intake) and biomarkers of effect on the body that could even be indicative of health/disease status (Picó et al., 2019).

It is widespread the use of *in vitro* assays or *in vivo* animal models on the research of mechanism of action of specific nutrients, food or diets; as the bias can be reduced due to the possibility of controlling environmental factors

as well as their genetical homogeneity. Moreover, animal models make possible to analyse various biofluids and tissues, some of them not feasible in humans trials (Braconi et al., 2018; Cifuentes, 2013).

Among the diseases whereby the intake of certain foods has an effect on their status, there are some in which their development is intrinsically related to diet. The so-called diet-related diseases include a great variety of diseases and disorders that affect different organs and systems (diabetes, cardiovascular and cerebrovascular diseases, diverse cancers, metabolic syndrome and obesity, etc); and they compose one of the major causes of mortality worldwide.

Liver steatosis or non-alcoholic fatty liver disease (NAFLD) is a high incidence metabolic disease characterised by fat accumulation in the liver that affects 20-30 % of world population, with high probability to progress to more severe diseases as steatohepatitis, cirrhosis or even cancer (Masarone et al., 2021). Hence the interest in using untargeted metabolomics is double, for phenotyping this disease as well as for developing intervention nutritional studies aimed to generate an improvement in health status and delay the disease progression.

In the present chapter, untargeted metabolomics was applied to study the effects of pterostilbene and resveratrol supplementation to rats with induced liver steatosis (**Scientific article IV**). This work arose in collaboration with the research group from the department of nutrition and food science from University of the Basque Country (Spain). In the research article by Gómez-Zorita et al. (2020) they demonstrated the similar prevention and reduction effects of pterostilbene and resveratrol on oxidative stress and inflammation in liver steatosis (Gómez-Zorita et al., 2020). This consecutive study uses the advantages offered by IMS-HRMS in untargeted metabolomics to analyse the differences induced by the treatments with both compounds in the liver metabolome

4.2 Scientific article IV

Food Chemistry (2021) Submitted

The potential of ion mobility separation in combination with high resolution mass spectrometry for the identification of biomarkers highlighted by untargeted metabolomics: the effects of pterostilbene and resveratrol consumption in liver steatosis, animal model

Leticia Lacalle-Bergeron[a], David Izquierdo-Sandoval[a], Alfredo Fernández-Quintela[b,c,d], María P Portillo [b,c,d]*, Juan V Sancho[a], Félix Hernández Herández[a], Tania Portolés[a]*

[a] Enviromental and Public Health Analytical Chemistry, Research Institute for Pesticides and Water (IUPA), University Jaume I, Av. Sos Baynat S/N, 12071 Castellón de la Plana, Spain

[b] Nutrition and Obesity group, Department of Nutrition and Food Science, Faculty of Pharmacy, University of the Basque Country (UPV/EHU), Lucio Lascaray Research Centre, 01006 Vitoria-Gasteiz, Spain

[c] BIOARABA Institute of Health, 01009 Vitoria-Gasteiz, Spain

[d] CIBERobn Physiopathology of Obesity and Nutrition, Institute of Health Carlos III (ISCIII), 01006 Vitoria-Gasteiz, Spain

Abstract

Untargeted metabolomics approach with the novel combination of ion mobility separation coupled to high resolution mass spectrometry (IMS-HRMS) have been applied in order to determine the impact of resveratrol and pterostilbene treatment on the metabolic profile of the Wistar rats liver with induced liver steatosis. Having studied previously the effect that resveratrol and pterostilbene had on the development of liver steatosis, the aim of this study is to determine the metabolic changes that occur in rat liver with diet-induced NAFLD (Non-Alcoholic Fatty Liver Disease) by administration of these compounds. RP-LC and HILIC in both ionisation modes were employed to analyse the liver samples (n = 40) from Wistar rats fed with four experimental high-fat and high-fructose diet (HF), supplemented or not with resveratrol (RSV30, 30 mg/kg/day) or pterostilbene (PT15 or PT30, 15 or 30 mg/kg/day, respectively). After univariate and multivariate statistical analysis, a total of 34 endogenous compounds present in the liver were elucidated as markers of the different administrated diets. Despite the similarity in the chemical structure between the two phenolic compounds under study, significantly different results were obtained regarding the changes induced in liver metabolism with NAFLD by their supplementation. Samples from resveratrol treatment revealed an alteration in phospholipid metabolism with 17 lysophospholipids highlighted as markers, while rats treated with pterostilbene showed a differentiating in vitamins and derivatives, among others. The IMS has demonstrated to be an improvement in the elucidation process and it has allowed the measurement of collision cross section value (CCS) as additional structural descriptor which could be compared with available libraries and provide more confidence in the identification.

Keywords

Untargeted Metabolomics, Ion Mobility, HRMS, Resveratrol, Pterostilbene, Liver Steatosis

1. Introduction

Foodomics is a relatively new discipline that has arisen as a result of the application of advanced analytical techniques (omics tools) and bioinformatics to nutritional and food research (Cifuentes, 2013). Among the omics tools, untargeted metabolomics (or metabolomic fingerprinting) approaches aim to compare patterns or fingerprints of metabolites that change a biological system or state in response to endogenous (genetics, disease…) or exogenous (environment, diet…) phenomena or condition (Lacalle-Bergeron et al., 2021).

As a result, this methodology can be a useful tool to seek new biomarkers in many working areas. Thus, it can provide biomarkers of diagnosis in several diseases highly prevalent in our society such hepatic steatosis (Karu et al., 2018), characterised by the accumulation of triglycerides in the liver (Madatali Abuwani et al., 2021), biomarkers of prognosis to identify the evolution of several diseases (Wang et al., 2011), such as the evolution of liver steatosis to steatohepatitis or cirrhosis (Lewinska et al., 2021; Pirola & Sookoian, 2018) and biomarkers to characterize the effects of treatments (i.e. diet, drugs,).

Regarding treatments, an emerging working area is the study of bioactive compounds with beneficial effects on health, which are either naturally present in food stuffs or artificially added, as in the case of "functional foods". In this context, the beneficial properties of phenolic compounds, present in fruits and vegetables, have been extensively studied. Thus, resveratrol (3,5,4′-trihydroxystilbene) is one of the most studied natural polyphenols with reported antioxidant and anti-inflammatory effects (Gimeno-Mallench et al., 2019). Nevertheless, it shows a low bioavailability due to the strong phase II metabolism that suffers. In turn, pterostilbene is a dimethoxy derivative of resveratrol, also showing antioxidant and anti-inflammatory effects; but higher bioability (Kapetanovic et al., 2011; Koh et al., 2021).

A key part of untargeted metabolomics is the characterisation of the wide variety of compounds that may be involved in the subject of study in complex and diverse biological matrices (Wolfender et al., 2015). This issue

can only be addressed with sufficiently sensitive and selective instruments, such as the hyphenation of chromatographic separations with high resolution mass spectrometry. The availability of compound libraries for metabolite identification, and tools and knowledge for data analysis and interpretation are also necessary (Vivanco et al., 2011).

Liquid chromatography coupled to high resolution mass spectrometry (LC-HRMS) in combination with untargeted metabolomics approaches has been widely employed in nutrition area, for instance to reveal the metabolic changes caused by the consumption of certain foods or bioactive compounds (Fu et al., 2022; Lacalle-Bergeron et al., 2020). On the one hand, liquid chromatography (LC) allows the use of a wide range of separation mechanisms because of the large variety of stationary phases available. In addition, it is more adequate for biological matrices due to their aqueous composition, requiring a less complex sample preparation for analysis than other chromatographic techniques such as gas chromatography (Segers et al., 2019). On the other hand, high resolution mass spectrometry (HRMS) is the most suitable option for the detection and identification of small metabolites owing to its high sensitivity and selectivity as well as the acquisition of accurate-mass full-spectrum data (Castro-Puyana et al., 2017). Moreover, the continuous improvements in instrumentation have allowed the incorporation of ion mobility spectrometry (IMS) to HRMS instruments, providing an additional structural information and enhancing the characterisation of the biomarkers (Segers et al., 2019). This technique provides the collision cross section value (CCS, $Å^2$) of each ion, an additional characterisation parameter, based on the measurement of drift time (DT), which is dependent of the individual size, shape and charge of each ion (Mairinger et al., 2018). In addition, IMS cell layout prior to hybrid HRMS analysers allows High Definition MS^E acquisition ($HDMS^E$). In conventional MS^E, where accurate-mass full-spectrum at low collision energy (LE) and high collision energy (HE) are acquired sequentially, information about the precursor ion and product ions is provided in a single injection. The incorporation of IMS data opens up the possibility to obtain cleaner fragmentation spectra without co-eluting ion fragments by filtering with the DT, as it is recorded for the

precursor ion of the LE spectra and linked to its product ions in the HE spectra (Paglia & Astarita, 2017).

The present work's aim was to apply an untargeted metabolomics approach, using UHPLC-IMS-HRMS, to perform a comparative study of the effects of pterostilbene and resveratrol on liver metabolome in rats showing liver steatosis induced by a diet rich in saturated fat and fructose, as well as to explore the capabilities that the IMS brings to conventional LC-HRMS in the identification of biomarkers.

2. Materials and methods

2.1. Chemicals and reagents

The solvents methanol and acetonitrile at LC-MS grade were purchased from Scharlab (Barcelona, Spain), as well as the eluent additive formic acid for LC-MS and reagent grade ammonium acetate. HPLC-grade water was obtained with Milli-Q water purification system (Millipore Ltd., Bedford, MA, USA). Leucine-enkephalin HPLC-grade (mass-axis recalibration) and analytical standards of riboflavin, cytidine, 1-methylnicotinamide, xanthosine, asymmetric dimethylarginine and docosahexaenoic acid were purchased werewas purchased from Sigma-Aldrich (Saint Louis, MO, USA).

2.2. Animals and study design

The experiments and animal protocols were approved by the Ethical Committee of University of the Basque Country (document reference M20_2015_245 CUEID), following the European regulations (European Convention-Strasburg 1986, Directive 2003/65/EC and Recommendation 2007/526/EC).

The study design is extensively described in Gómez-Zorita et al., 2020 (Gómez-Zorita et al., 2020). For this purpose, fifty male Wistar rats (6-week-old, 140-150 g) were housed in pairs in polycarbonate cages, after a 6-day adaptation perior, under controlled conditions of temperature (22 ± 2 °C) and 12 h light-dark cycle. The rats were randomly assigned to five groups (n = 10/group): the control group (CC) was fed with a healthy balanced diet (commercial standard diet AIN-93G, OpenSource Diets, Gentofte, Denmark, D10012G); the high-fat and high-fructose group (HF) was fed with a diet

containing 40 % of lipids and 22 % of fructose (OpenSource Diets, Gentofte, Denmark, D09100301); the PT15 and PT30 groups were fed with the same high-fat and high-fructose diet supplemented with pterostilbene in the amounts needed to provide doses of 15 mg/kg body weight/day (PT15) or 30 mg/kg body weight/day (PT30), and the RSV30 group was fed with the same high-fat and high-sucrose diet supplemented with resveratrol in the amount needed to provide 30 mg/kg body weight/day. During the experiment, all animals had *ad libitum* access to food and water. Food intake and body weight were recorded on a daily basis.

After the 8 weeks of experimental period, the animals underwent 12-hour fasting and sacrificed by cardiac exsanguination under anaesthesia (chloral hydrate).

2.3. Liver sampling and sample treatment

After exsanguination, livers were weighted and dissected into lobes that were individually stored at -80 °C until analysis. For this analysis the same lobe of each liver was used.

For sample treatment, 0.5 mL of cold water:methanol (1:1) was added to ~100 mg of liver sample and triturated. Then, 0.5 mL of cold water:methanol (1:1) were added again and mixed with a Vortex for 45 min. After another 45 min in a cold ultrasound bath and centrifuging at 12,000×rpm and a radius of 5.5 cm, 8855 g (RCF), for 30 min at 4 °C, the supernatant was divided into 3 aliquots: two vials of 300 µL were stored at −30 °C, one vial of 200 µL was stored at −80 °C. Moreover, the Quality Control sample (QC) was generated by pooling and mixing 50 µL of each sample extract.

Samples were randomly injected into the UHPLC-IMS-QTOF MS system in order to compensate/reduce potential instrumental drift. Assuming that QC is a representative average sample formed by a pool equivalent aliquot of all final sample extracts, it was used for both column stabilisation purposes (by 10 QC injections at the beginning of each sample batch for RP-LC and 15 for HILIC) and, following by the injection every 10 samples, to control possible instrumental drift throughout the sequence

2.4. Instrumentation

Samples were analysed using ultra-high performance liquid chromatography (UHPLC) with a Waters ACQUITY UHPLC I-Class system (Waters, Milford, MA, USA) coupled to a VION® IMS QTof (Waters, Manchester, UK), using an electrospray ionisation interface operating in both positive (ESI+) and negative (ESI−) mode. Equipment control and data acquisition and processing were performed using UNIFI software (V.1.9.2, Waters, Manchester, UK).

Two chromatographic separations were employed to cover a wider range of compound polarities. Reversed Phase Liquid chromatography (RP-LC) was used with a CORTECS® C18 fused-core 2.7 μm particle size analytical column 100 x 2.1 mm (Waters), whereas a CORTECS® HILIC fused-core 2.7 μm particle size analytical column 100 x 2.1 mm (Waters) was employed for Hydrophilic Interaction Liquid Chromatography (HILIC). For both types of liquid chromatography and both ionisation modes, gradients elution were performed at 0.3 mL/min flow rate, 40 °C column oven temperature and 1 μL sample injection volume were selected for.

The RP-LC gradient elution was performed using water (A) and methanol (B) as mobile phases, both with 0.01 % of formic acid, changing as follows: 10 % B at 0 min to 90 % B at 14 min, 90 % B at 16 min, and 10 % B at 16.01 min, with a total run time of 18 min. The same gradient was employed for both ionisation modes.

For HILIC separation, mobile phases acetonitrile:water (95:5, *v/v*) (A) and water (B), both with 0.01 % formic acid and 10 mM ammonium acetate, were employed. The gradient started with 2 % B until 1 min, 60 % B at 10 min, 60 % B at 12 min and finally 2 % B at 12.01 min, with a total run time of 15 min. This gradient was the same for both ionisation modes.

The ESI operated at a capillary voltage of 0.7 kV and 2.00 kV for positive (ESI+) and negative (ESI−) electrospray ionisation mode, respectively. In both cases the cone voltage was 30 V, source temperature was set to 120 °C and desolvation gas to 550 °C with a flow rate of 1000 L/h. Nitrogen was used as the desolvation gas, nebulizing gas, mobility gas and collision gas. The mass spectrometer was operated in ion mobility ($HDMS^E$) mode for acquisition in both polarities over an *m/z* range of 50–1000 Da and

a scan time of 0.3 s. In HDMS^E experiments, two acquisition functions were acquired simultaneously: low-energy function (LE), with a fixed collision energy of 6 eV, and high-energy function (HE) with a collision energy ramp from 28 to 56 eV.

Calibrations of mass axis and DT were performed monthly with the "Major Mix IMS/T of calibration kit" supplied by the vendor (Waters) and Leucine-Enkephalin solution (100 µg/L acetonitrile:water (50:50, *v/v*) containing 0.01% formic acid) was employed for continuous recalibration of the mass axis and ensure the robust accurate mass measurement along chromatographic runs.

2.5. Data processing and statistical analysis

The raw data (*.uep*, UNIFI, Waters) were imported to Progenesis QI (V.2.5, Nonlinear Dynamics, Newcastle, UK) for baseline filter, peak alignment and other data analysis. The software automatically performs 4D peak picking (based on the intensity, *m/z*, retention time and DT), retention time alignment using QC replicates as reference (except for the first 9 QC injections used for column stabilisation in RP-LC or first 14 QC injection for HILIC); and response normalisation. The peak picking conditions were set as follows: all runs, limits (automatic), sensitivity (automatic, level 2), chromatographic peak width (minimum peak width of 0.1 min), and retention time limits (0.3 to 17 min and 0.3 to 12 min, for RP and HILIC respectively). To apply the deconvolution tool, the selected adducts ions forms $[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, $[M-H_2O+H]^+$, $[2M+H]^+$ and $[2M+Na]^+$ were selected for positive ionisation analysis; and $[M-H]^-$, $[M-H_2O-H]^-$, $[M+Cl]^-$, $[M+FA-H]^-$ and $[2M-H]^-$ for negative ionisation analysis. Samples were originally divided into 5 groups (CC, HF, RSV, PT15 and PT30) in the "Experimental Design Setup", following the "Between-subject Design" comparison (samples from a given subject appear in only one condition). The software will then perform a One-way ANOVA calculation assuming the independence of each sample followed by a false discovery rate (FDR) optimisation approach; to test the differences among the experimental groups. The levels of statistical significance were set at 95 % level (*p-value* ≤ 0.05).

The processed data were then directly exported to EZinfo (V.3.03, Umetrics, Sweden) for multivariate statistical analysis. Firstly, Principal Component Analysis (PCA), an unsupervised analysis, was applied to ensure the correct grouping of the QC replicates samples in the centre of the plot after normalisation and the absence of outliers. Then, Partial Least Square–Discriminant Analysis (PLS-DA) was conducted to maximize the separation between the groups and the validation of the model was performed by leaving-1/7-out cross-validation approach. Ultimately, an Orthogonal PLS-DA (OPLS-DA) was carried out to highlight the most robust markers (threshold $p(corr) \geq |0.6|$ and $p[1]\ loading \geq |0.1|$).

2.6. Elucidation workflow

The most significant markers highlighted in the OPLS-DA were tentatively elucidated based on their accurate masses, DT filtered HE spectra information and CCS ($Å^2$). The search in reliable mass spectra databases as Metlin, HMDB and Lipid Maps allowed the annotation of metabolites, comparing the HE spectra to the available ones or to in-silico fragmentation spectra. In addition, CCS libraries were also employed. The final identity could only be confirmed by comparing the retention time, fragmentation and CCS with a commercially available standard. When not available, CCS values were predicted by means of our CCS prediction model (Bijlsma et al., 2017) (Celma et al. 2021 Analytical Chemistry submitted) aimed at providing additional identification confidence. Different levels of confidence in metabolite identification were employed according to Schymanski et al. (2014) classification (Schymanski et al., 2014) updated including ion mobility separation as an additional parameter for more reliable identification in Celma et al. (2020) (Celma et al., 2020).

3. Results and discussion

3.1. Sample treatment

Regarding the sample treatment, in comparison to fluid biological sample matrices, a special care has to be taken in the pre-treatment, homogenisation and metabolite extraction stages from a tissue. Firstly, to avoid metabolome changes due to temperature and enzymatic activity, more probable in tissues, the storage and transport of liver samples were kept at -

80 ˚C until analysis. It was intended to perform all the sample treatment under low temperatures conditions as possible, using previously cooled solvents and employing instrumentation able to work under low temperature condition. Especially larger stages of vortex and ultrasound sonication, in comparison to biological fluids as plasma or saliva, were employed to maximise the homogenisation and to perform and exhaustive solvent-extraction of the metabolites.

A mixture of water:methanol (50:50, *v/v*) allowed to maximize the polarity coverage of the metabolites and the use of methanol in the mixture eliminated macromolecules (nucleic acids and proteins) present in the sample by deproteinisation and, thereby, focus on the low-weight molecules (metabolites). After the centrifugation to eliminate the solid residues and precipitated proteins, the resulting extract was directly injected in both chromatographic separations (RP and HILIC) without the need for additional evaporation and redissolution in more suitable solvents. This was possible due to the low injection volume needed (1 µL) as a consequence of the high sensitivity of the UHPLC-IMS-QTOF MS instrument. Two different LC separations were employed to detect as many compounds as possible, in positive (pos) and negative (neg) ionisation modes. RP (with a C18 column) is suited for nonpolar hydrophobic metabolites and HILIC (silica based column) for polar hydrophilic compounds. Therefore, four sample batches were performed: RP pos, RP neg, HILIC pos and HILIC neg. To avoid bias during the analysis the sample extracts were randomly injected and QC samples were injected every 10 samples for instrumental drift control and normalisation.

The data were all acquired in HDMS^E, where the ion mobility DT measurements are combined with the information of the LE and HE spectra acquired simultaneously, obtaining information from the precursor ion and full-scan accurate mass fragmentation, respectively. Therefore, a 4D data matrix was obtained at the end of the analysis, allowing to characterize each feature by the retention time from the chromatographic separation, the CCS calculated from the DT, the accurate mass and fragmentation spectra; and intensity. As the ion mobility separation is prior to the fragmentation, the "precursor" ion shares the same DT than its fragments on the three-

dimensional plots, allowing to align the feature with its fragments and reduce the interference of co-eluting "precursor" ions. Therefore, DT is especially useful not only as characterisation parameter but also to obtain cleaner HE spectra of almost MS/MS quality as it allows to only shows the fragments that have been generated from a "precursor" ion with a given DT, and therefore, enhancing the structural elucidation.

3.2. Data processing and statistical analysis

The data were acquired with UNIFI software (waters, UK) and the raw data were exported to **.uep* data format (UNIFI export package) to be then processed by Progenesis QI (Nonlinear Dynamics, UK). To the best of our knowledge, it is the only program able to interpret and process the four-dimensional data (RT, *m/z*, area and CCS) for *–omics* purposes. After the data import, processing with Progenesis QI starts with retention time alignment, peak picking and normalisation. All the samples were successfully aligned with a score greater than 85 % and the normalisation was performed using all compounds. The peak picking of the four data sets RP pos, RP neg, HILIC pos and HILIC neg resulted in the detection of 4683, 2800, 4231 and 2135 features, respectively. Those data set were reduced to 3298, 2468, 2439 and 1437 respectively, after the removal of compounds that exhibit a poor stability with a relative standard deviation (% RSD) above 30 % within the QCs. The deconvolution tool allowed to group the features coming from the same compound, according to the adducts specified in the processing, and annotate them under the same label (*xx.xx_yyy.yyyyn*, *xx.xx* being the retention time in minutes and *yyy.yyyy* the exact neutral mass when more than one adduct ion is found for the same compound; or *xx.xx_zzz.zzzzzm/z*, *zzz.zzzz* being the exact mass of the single ion found).

Multivariate unsupervised analysis PCA was then applied to each data set as exploratory visualisation of the results in order to observe trends, groupings and/or outliers. Firstly, QC replicates injected after every 10 samples (n = 9 per data set) should be clustering in the centre of the PCA score plots. Because they are a pool of all analysed samples extracts, they should act as an "average" sample and thus demonstrate by highlighting that the difference between the groups are not caused by instrumental drift, and therefore, the correct acquisition of the data. **Figure 1** shows the PCA score

plot obtained for the four data sets, where the correct clustering of QC replicates is evident in all of them, proving the proper performance of the analytical system along all, four runs. As it can be observed there is also a clear differentiation of the liver samples of rats that had been given a standard diet (CC) from those that were fed with high-fat high-fructose diet, supplemented or not with the compounds under study. This is a predictable result given that these samples were already differentiated at naked eye in the sample treatment.

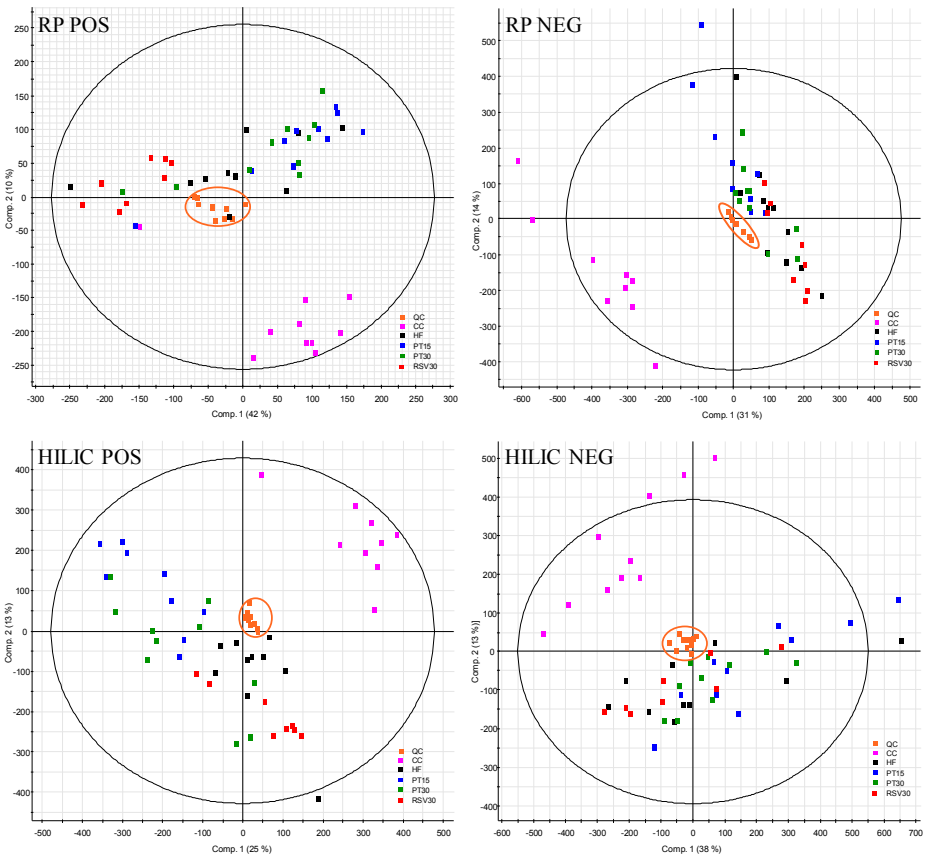


Figure. 1. *PCA score plot component 1 vs component 2 for liver metabolic profiles in the four different analysis modes. The QC samples (QC ■) are grouped and centred in the plot. Five groups of liver samples from Wistar rats fed with different diets were analysed: standard diet (CC ■), high-fat and high-fructose diet (HF ■), and high-fat and high-fructose diet enhanced with 15 mg/kg/day or 30 mg/kg/day of pterostilbene (PT15 ■ and PT30 ■) or 30 mg/kg/day of resveratrol (RSV30 ■). The variance explained 79 %, 89 %, 75 % and 80 % for RP pos, RP neg, HILIC pos and HILIC neg, respectively.*

Given that the objective of this study was the metabolic differentiation between the livers of NAFLD-induced rats whose diets have been supplemented with pterostilbene and resveratrol, the group of CC samples were discarded for the following statistical analysis. In addition, it can be observed that PCA does not manage to clearly differentiate between resveratrol and pterostilbene supplementation from the non-supplemented samples (HF), and that the inherent differences of these samples are still too great to approach the healthy liver samples. With the 2282, 1410, 1237 and 550 statistically significant features for RP pos, RP neg, HILIC pos and HILIC neg, respectively, a second PCA analysis was employed to interrogate the data. **Figure S1** shows the PCA score plot of the first two components obtained comparing the four groups HF, PT15, PT30 and RSV30. As it can be observed, the inherent differentiation obtained by this unsupervised analysis seems to be not significant enough to clearly separate the groups, although it can be seen already a trend in the separation of RSV30 group from the others and the positioning of the HF group approximately in the centre of the graph. The presence of outliers was studied based on the 95 % of confidence limit of the Hotelling's T2 Range.

Then, supervised multivariate statistical PLS-DA modelling was applied to highlight the differences between the experimental groups based on the statistically significant features. **Figure S2** shows the PLS-DA score plot in the plane of the first two latent variables for the four data sets. The separations achieved were not ideal, obtaining coefficients for variance explained (R^2Y) from 36 to 51 % and coefficients for variance predicted (Q^2) from 18 to 40 % (**Table 1**). These results are close and even lower than the accepted limit of 50 % for variance explained and 40 % for predicted for biologic al models (Worley & Powers, 2012). This is primarily due to the low differentiation between the two dosages of pterostilbene (PT15 and PT30), demonstrating that the different dosages concentrations with pterostilbene apparently do not imply a significant variation at metabolic level. Then, it was attempted the PLS-DA differentiation considering the two dosages of pterostilbene as a single group. For the discrimination between the three resultant groups, 2457, 1388, 1325 and 557 statistically significant features were obtained for RP pos, RP neg, HILIC pos and HILIC neg, respectively.

Figure 2 shows the PLS-DA score plots obtained for the four data sets, where the differentiation has been considerably improved, obtaining R^2Y from 70 to 86 %, and Q^2 from 54 to 66 % (**Table 1**), resulting in 95 %, 89 %, 84 % and 84 % of the samples classified correctly for RP pos, RP neg, HILIC pos and HILIC neg, respectively. In all data sets, the first latent variable separates the two supplemented diets with pterostilbene or resveratrol (PT and RSV30); while the second component differentiated the supplemented diets from the HF group.

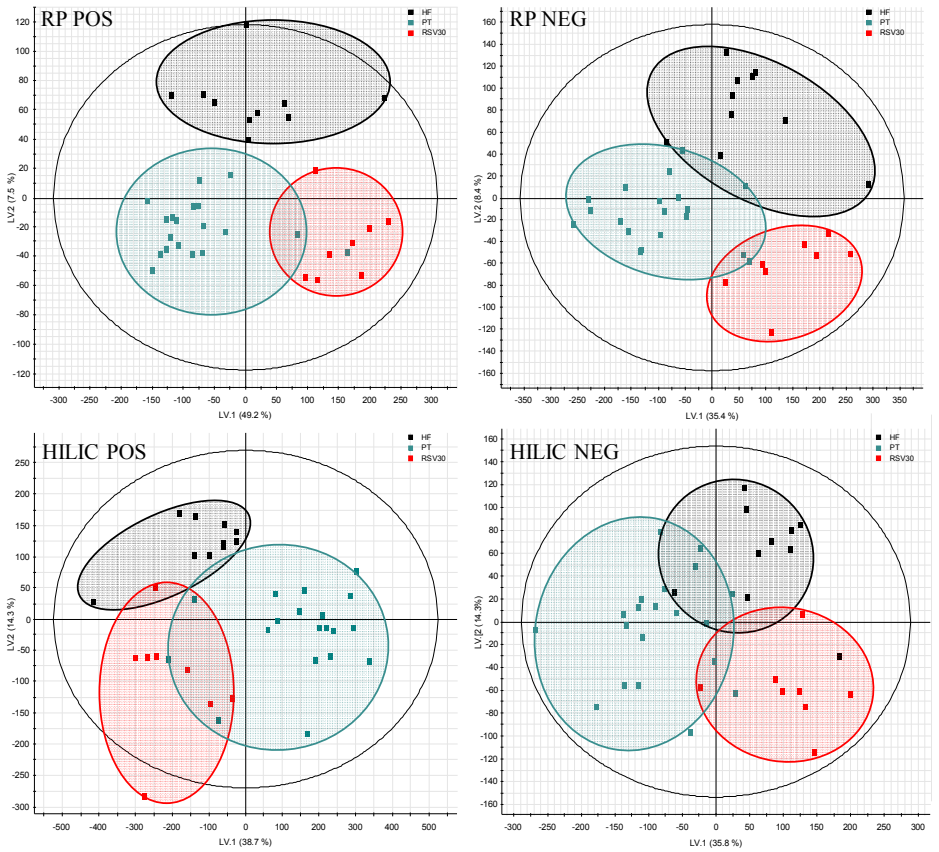


Figure. 2. *PLS-DA score plot LV1 vs LV2 for liver metabolic profiles in the four different analysis modes. Three groups of liver samples from Wistar rats fed with different diets were analysed: high-fat and high-fructose diet (HF ■), and high-fat and high-fructose diet enhanced with 15 mg/kg/day or 30 mg/kg/day of pterostilbene (PT ■) or 30 mg/kg/day of resveratrol (RSV30 ■). Being the coefficients R^2 for the variance explained and Q^2 for the variance predicted, it was obtained: R^2Y = 86 % and Q^2 = 66 % for RP pos, R^2Y= 79 % and Q^2 = 55 % for RP neg, R^2Y = 70 % and Q^2 = 54 % for HILIC pos and R^2Y = 78 % and Q^2 = 60 % for HILIC neg.*

In order to highlight the primary features for the discrimination between the groups, an Orthogonal PLS-DA (OPLS-DA) was finally employed, where only two groups can be faced. Therefore, to highlight the most significant features three classifications were attempted by this method: HF *vs* RSV, HF *vs* PT and RSV *vs* PT. For all the facing, a variance explained (R^2Y) above 78 % and a variance predicted (Q^2) above 74 % were obtained, except for PT *vs* RSV facing in HILICpos where it was founded a Q^2 = 58 % (**Table 2**). S-Plots were generated from each OPLS-DA , allowing an easiest visualisation of the markers with higher discrimination power between the two-faced groups. From the S-plot, the most relevant ones were the features closer to *p(corr)* 1 or -1. To select the most relevant features a cut-off *p(corr)* ≥ |0.6| and *p[1] loadings* ≥ |0.1| were employed, obtaining a sum up of 117 features between the four data sets. Nevertheless, the list of possible markers was reduced still further to 34 for different reasons. For example, some of the compounds were detected more than once with the different chromatographic separation or even different polarities (e.g. feature HILICpos_2.05_376.1390n and RPpos_4.00_376.1388n both with the same CCS of 186 $Å^2$, in addition, this compound was also present in RP neg as RPneg_3.99_376.1386n). In other cases, some types of adducts or in-source fragments had not been specified in the deconvolution step of Progenesis QI, and therefore they had not been clustered and appeared as an independent feature. Finally, those features with poor signal or bad peak shape that cast doubt on their validity as markers were discarded.

Table 1. *Parameters of the PLS-DA models.*

Statistical model / Characteristics	***PLS-DA model diagnostics***							
	RP pos		RP neg		HILIC pos		HILIC neg	
Groups	4 groups: PT15, PT30, RSV and HF	3 groups: PT, RSV and HF	4 groups: PT15, PT30, RSV and HF	3 groups: PT, RSV and HF	4 groups: PT15, PT30, RSV and HF	3 groups: PT, RSV and HF	4 groups: PT15, PT30, RSV and HF	3 groups: PT, RSV and HF
Components	2	4	3	4	2	3	2	4
Variance explained – R^2Y (cum)	48 %	86 %	51 %	79 %	45 %	70 %	36 %	78 %
Variance predicted – Q^2 (cum)	41 %	66 %	33 %	55 %	37 %	54 %	18 %	60 %

R^2 - fit how model fits the data and Q^2 – predictive ability, by seven-round internal cross-validation as default option of EZinfo software (Umetrics, Sweden)

Table 2. *Parameters of the OPLS-DA models.*

Statistical model / Characteristics	***OPLS-DA model diagnostics***											
	RP pos			RP neg			HILIC pos			HILIC neg		
Groups	PT vs HF	RSV vs HF	PT vs RSV	PT vs HF	RSV vs HF	PT vs RSV	PT vs HF	RSV vs HF	PT vs RSV	PT vs HF	RSV vs HF	PT vs RSV
	n =20 vs 10	*n* =10 vs 10	*n* =20 vs 10	*n* =20 vs 10	*n* =10 vs 10	*n* =20 vs 10	*n* =20 vs 10	*n* =10 vs 10	*n* =20 vs 10	*n* =20 vs 10	*n* =10 vs 10	*n* =20 vs 10
Components	4	6	4	6	4	6	2	4	2	3	3	2
Goodness-of-fit parameter - R^2X	77 %	90 %	77 %	90 %	77 %	90 %	60 %	87 %	59 %	89 %	64 %	54 %
Variance explained – R^2Y (cum)	96 %	99 %	96 %	99 %	96 %	99 %	86 %	99 %	78 %	88 %	99 %	83 %
Variance predicted – Q^2 (cum)	87 %	88 %	87 %	88 %	87 %	88 %	79 %	97 %	58 %	79 %	94 %	74 %

R^2 - fit how model fits the data and Q^2 – predictive ability, by seven-round internal cross-validation as default option of EZinfo software (Umetrics, Sweden)

3.3. Elucidation Process

A total of 34 markers were selected for further identification. Experimental data are recorded in **Table S1** and statistical relevance is shown in **Table 3** and **Figure 4**, **5** and **6**. In order to accomplish the elucidation of those compounds, the first step was to annotate the candidates thanks to Progenesis QI identification tool which performs a combination of neutral mass, MS/MS and CCS (if available) based searches by comparison with the HDMS^E spectra and CCS data of each putative marker. The annotation of each marker was carefully reviewed based on mass accuracy and both parents and fragment ions and CCS from UNIFI raw data. Following the criteria of our laboratory and the identification level system described by Celma et al. (2020) (Celma et al., 2020), the different identification levels were given to the markers according to the data available for each of them.

Table 3. *Statistical values for the 34 markers tentatively identified.*

Nº	Compound	Elemental composition	*p(corr)*	Marker of	Chromatography	ESI
1	Riboflavin (Vitamin B2)	$C_{17}H_{20}N_4O_6$	0.85	PT, in HF vs PT	HILIC	pos
			-0.74	PT, in RSV vs PT		
			0.73	PT, in HF vs PT	RP	pos
			0.74	PT, in HF vs PT	RP	neg
2	Allopurinol	$C_5H_4N_4O$	0.77	PT, in HF vs PT	HILIC	pos
3	Cytidine	$C_9H_{13}N_3O_5$	-0.75	HF, in HF vs RSV	HILIC	pos
4	Methylnicotinamide	$C_7H_8N_2O$	-0.67	HF, in HF vs PT	HILIC	pos
			-0.62	HF, in HF vs RSV		
5	Nicotinamide Mononucleotide	$C_{11}H_{15}N_2O_8P$	0.70	PT, in HF vs PT	HILIC	pos
			0.67	RSV, in HF vs RSV		
6	Xanthosine	$C_{10}H_{12}N_4O_6$	-0.70	PT, in PT vs RSV	RP	neg
7	Indoleacrylic acid	$C_{11}H_9NO_2$	-0.61	HF, in HF vs PT	RP	pos
			-0.84	HF, in HF vs RSV		
8	4,5-didehydro-5-deoxyadenosine	$C_{10}H_{11}N_5O_3$	0.64	PT, in HF vs PT	HILIC	pos
			-0.65	PT, in RSV vs PT		
9	1-(a-ribofuranosyl)-lumichrome	$C_{17}H_{18}N_4O_6$	-0.76	HF, in HF vs RSV	HILIC	pos
10	Adenosine-5'-(O-methylphosphate)	$C_{11}H_{16}N_5O_7P$	0.62	RSV, in HF vs RSV	HILIC	pos
			0.73	RSV, in HF vs RSV	HILIC	neg
			0.86	PT, in HF vs PT		
			0.82	PT, in HF vs PT	RP	pos
			0.77	RSV, in HF vs RSV	RP	neg
			0.89	PT, in HF vs PT		

Table 3. *Statistical values for the 34 markers tentatively identified (continuation).*

Nº	Compound	Elemental composition	*p(corr)*	Marker of	Chromatography	ESI
11	Adenosylmethionine	$C_{15}H_{22}N_6O_5S$	0.67	PT, in HF vs PT	HILIC	pos
12	ADMA (asymmetric dimethylarginine)	$C_8H_{18}N_4O_2$	-0.77	HF, in HF vs PT	HILIC	pos
13	Oxidised Glutatione	$C_{20}H_{32}N_6O_{12}S_2$	0.70	PT, in HF vs PT	HILIC	pos
			-0.62	PT, in RSV vs PT		
			-0.83	PT, in RSV vs PT	HILIC	neg
			-0.70	PT, in RSV vs PT	RP	pos
			-0.92	PT, in RSV vs PT	RP	neg
14	Resveratrol-O-sulfate	$C_{14}H_{12}O_6S$	0.82	RSV, in HF vs RSV	HILIC	neg
			0.85	RSV, in RSV vs PT		
15	Pterostilbene-4'-O-sulfate	$C_{16}H_{16}O_6S$	0.63	PT, in HF vs PT	HILIC	neg
			-0.70	PT, in RSV vs PT		
16	Docosahexaenoic acid (DHA)	$C_{22}H_{32}O_2$	0.84	RSV, in HF vs RSV	HILIC	neg
			0.71	PT, in HF vs PT		
			0.82	RSV, in HF vs RSV	RP	pos
			0.92	RSV, in HF vs RSV	RP	neg
			0.64	PT, in HF vs PT		
17	Docosapentaenoic acid (DPA)	$C_{22}H_{34}O_2$	0.80	RSV, in HF vs RSV	HILIC	neg
			0.62	PT, in HF vs PT		
			0.85	RSV, in HF vs RSV	RP	neg
			0.66	PT, in HF vs PT		
18	LysoPE(16:0)	$C_{21}H_{44}NO_7P$	0.72	RSV, in HF vs RSV	RP	neg
19	LysoPE (18:0)	$C_{23}H_{48}NO_7P$	0.81	RSV, in HF vs RSV	RP	neg
20	LysoPE(18:1)	$C_{23}H_{46}NO_7P$	0.71	RSV, in HF vs RSV	HILIC	pos
			0.69	RSV, in HF vs RSV	HILIC	neg
21	LysoPE(18:2)	$C_{23}H_{44}NO_7P$	0.71	RSV, in HF vs RSV	HILIC	neg
22	LysoPE(20:4)	$C_{25}H_{44}NO_7P$	0.86	RSV, in HF vs RSV	HILIC	pos
			0.88	RSV, in PT vs RSV		
			0.69	RSV, in HF vs RSV	HILIC	neg
			0.69	RSV, in PT vs RSV	RP	neg
23	LysoPE(22:6)	$C_{27}H_{44}NO_7P$	0.88	RSV, in HF vs RSV	HILIC	pos
			0.72	RSV, in HF vs RSV	HILIC	neg
24	LysoPC(16:0)	$C_{24}H_{50}NO_7P$	0.84	RSV, in HF vs RSV	HILIC	pos
			0.92	RSV, in PT vs RSV		
			0.74	RSV, in HF vs RSV	RP	pos
25	LysoPC(18:1) isomer 1	$C_{26}H_{52}NO_7P$	0.88	RSV, in PT vs RSV	HILIC	pos

Table 3. *Statistical values for the 34 markers tentatively identified (continuation).*

Nº	Compound	Elemental composition	*p(corr)*	Marker of	Chromatography	ESI
26	LysoPC(18:1) isomer 2	$C_{26}H_{52}NO_7P$	0.82	RSV, in HF vs RSV	HILIC	pos
			0.92	RSV, in PT vs RSV		
27	LysoPC(18:3)	$C_{26}H_{48}NO_7P$	0.78	RSV, in HF vs RSV	HILIC	pos
			0.88	RSV, in PT vs RSV		
28	LysoPC(20:4) isomer 1	$C_{28}H_{50}NO_7P$	0.75	RSV, in PT vs RSV	HILIC	pos
29	LysoPC(20:4) isomer 2	$C_{28}H_{50}NO_7P$	0.72	RSV, in PT vs RSV	HILIC	pos
30	LysoPC(22:6)	$C_{30}H_{50}NO_7P$	0.74	RSV, in PT vs RSV	HILIC	pos
31	LysoPI(20:4)	$C_{29}H_{49}O_{12}P$	0.81	RSV, in HF vs RSV	HILIC	neg
32	LysoPS(20:4)	$C_{26}H_{44}NO_9P$	0.91	RSV, in HF vs RSV	HILIC	neg
33	LysoPS (21:0)	$C_{27}H_{54}NO_9P$	0.87	RSV, in HF vs RSV	RP	neg
34	LysoPS (2-OMe-18:0)	$C_{25}H_{52}NO_9P$	0.78	RSV, in HF vs RSV	RP	neg
			0.90	RSV, in HF vs RSV		

Marker 1 has been selected as an illustrative example of the elucidation process followed in the present work (**Figure 3**). This marker was found in HILIC pos (HILICpos_2.05_376.1390n) as marker of pterostilbene supplementation in PT vs HF (*p(corr)* 0.85) and PT *vs* RSV (*p(corr)* -0.74); in RP pos analysis (RPpos_4.00_376.1388n, marker of PT in PT *vs* HF, *p(corr)* 0.73); and also in RP neg analysis (RPneg_3.99_376.1386n, marker of PT in HF *vs* PT). This marker was also present in HILIC neg with a *p(corr)* of 0.54, below the threshold applied for the selection. Firstly, thanks to the fact that ion mobility separation takes place prior to the fragmentation, both the parent and the fragment ions have associated the same DT. Hence, it is possible obtaining of cleaner mass spectra without co-eluting/interfering ions when it is filtered by the DT of the parent ion, as it can be observed in **Figure 3B** and **3C** (DT filter 5.55 ± 0.22 ms, data form HILIC pos) in comparison with the unfiltered ones (**Figure 3D** and **3E**). Therefore, it is obtained a quasi-MS^2 spectrum with composite of "product" ions obtained at different collision energies (HE energy ramp from 28 to 56 eV) pure enough and with a lot of information, which would avoid the re-injection of the samples. The filtered spectra obtained from the analysis of the same ionisation modes were mostly identical (RP pos to HILIC pos and RP neg to HILIC neg). Nevertheless, the unfiltered spectra may give also useful

information. Since the adducts are formed in the ion source previous to the IMS separation, $[M+Na]^+$, $[M+K]^+$ and $[M-H_2O+H]^+$ adducts DT was different enough to the protonated molecule to be filtered by DT and they can only be observed in the unfiltered LE spectra (**Figure 3D**). These adducts were all successfully deconvoluted by Progenesis QI and assigned as the same compound.

The most likely elemental composition for this marker was found to be $C_{17}H_{20}N_4O_6$ (error: +0.2, -0.01 and +0.8 mDa in HILIC pos, RP pos and RP neg, respectively). The mean retention time across the samples in HILIC was 2.00 min and 4.00 min in RP, with a CCS of 186 $Å^2$ for the protonated and 190 $Å^2$ for the deprotonated molecule. **Figure S3** shows the LE and HE DT-filtered spectra from deprotonated molecule obtained in RP neg analysis where complementary information of the elucidated structure was obtained. A proposed structure fragmentation was raised for both ionisation modes, and it was supported with the experimental spectra from METLIN with a fragmentation match over 60 % in Riboflavin (vitamin B2). Moreover, due to the availability of experimental CCS databases, this compound obtained a match with the available *Metabolic Profiling CCS Library* for Progenesis QI, with a delta error of + 0.23 % (maximum CCS tolerance 2.5 %), increasing considerably the confidence in the Riboflavin identification.

Among the list of markers, related compounds showed similar fragmentation. Marker 9 was found in HILIC pos analysis (HILICpos_1.28_375.1304*m/z*) as marker of high-fructose high-fat diet (HF group) in RSV vs HF (*p(corr)* -0.76), obtaining a level 3 tentative elucidation as 1-(a-ribofuranosyl)-lumichrome with molecular formula $C_{17}H_{18}N_4O_6$ (see marker structure in **Figure S4**). This marker shared with Riboflavin the non-specific water loss (*m/z* 357.1182, -1.69 mDa), and fragments *m/z* 243.0882 ($[C_{12}H_{11}N_4O_2]^+$, -0.0 mDa), *m/z* 198.0664 ($[C_{11}H_8N_3O]^+$, -0.3 mDa) and *m/z* 172.0868 ($[C_{10}H_{10}N_3O]^+$, -0.7 mDa) observed in the DT filtered HE spectra of riboflavin (**Figure 3C**), which shows that they share in their structure the nitrogenous base flavin.

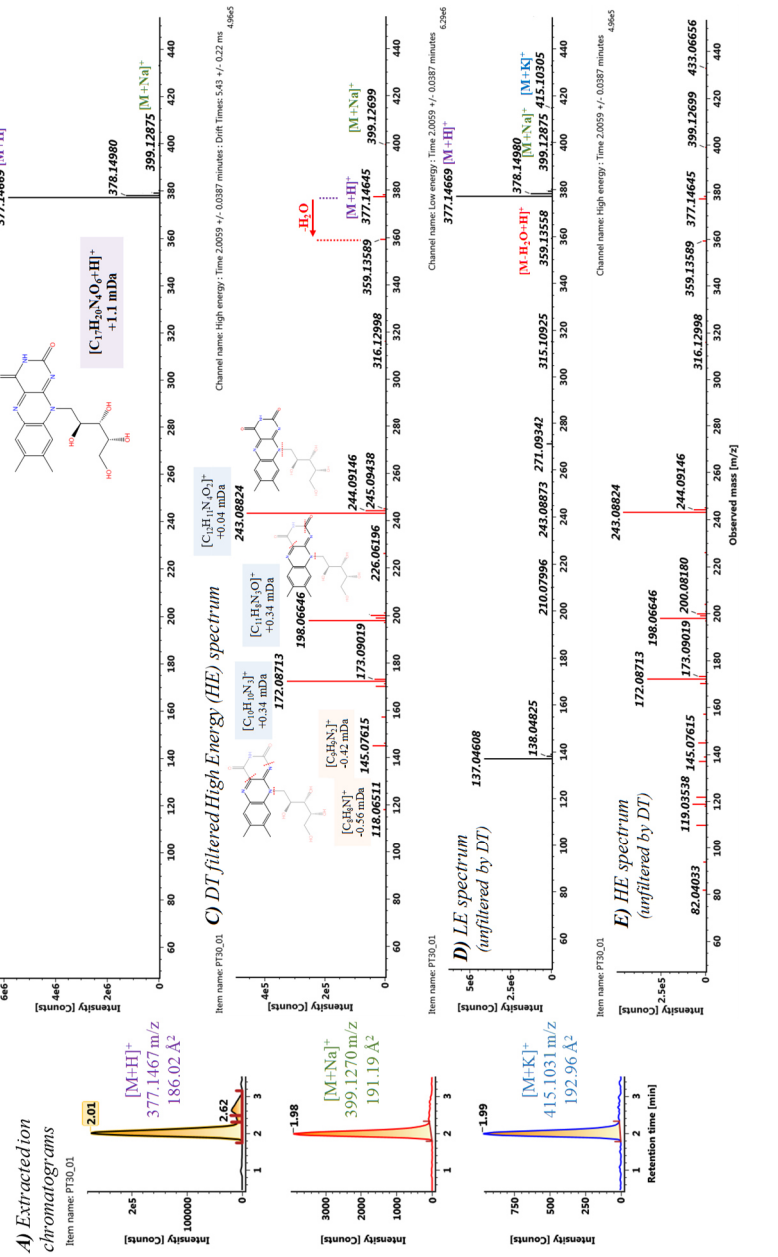


Figure. 3. *Elucidation of marker 1 Riboflavin (vitamin B2) based in chromatograms and HDMS^E spectra obtained from HILIC pos analysis:* ***A)*** *Extracted ion chromatograms of adducts found, along with the experimental m/z and CCS,* ***B)*** *DT filtered LE spectrum,* ***C)*** *DT filtered HE spectrum,* ***D)*** *LE spectrum and* ***E)*** *HE spectrum. A structure fragmentation is proposed.*

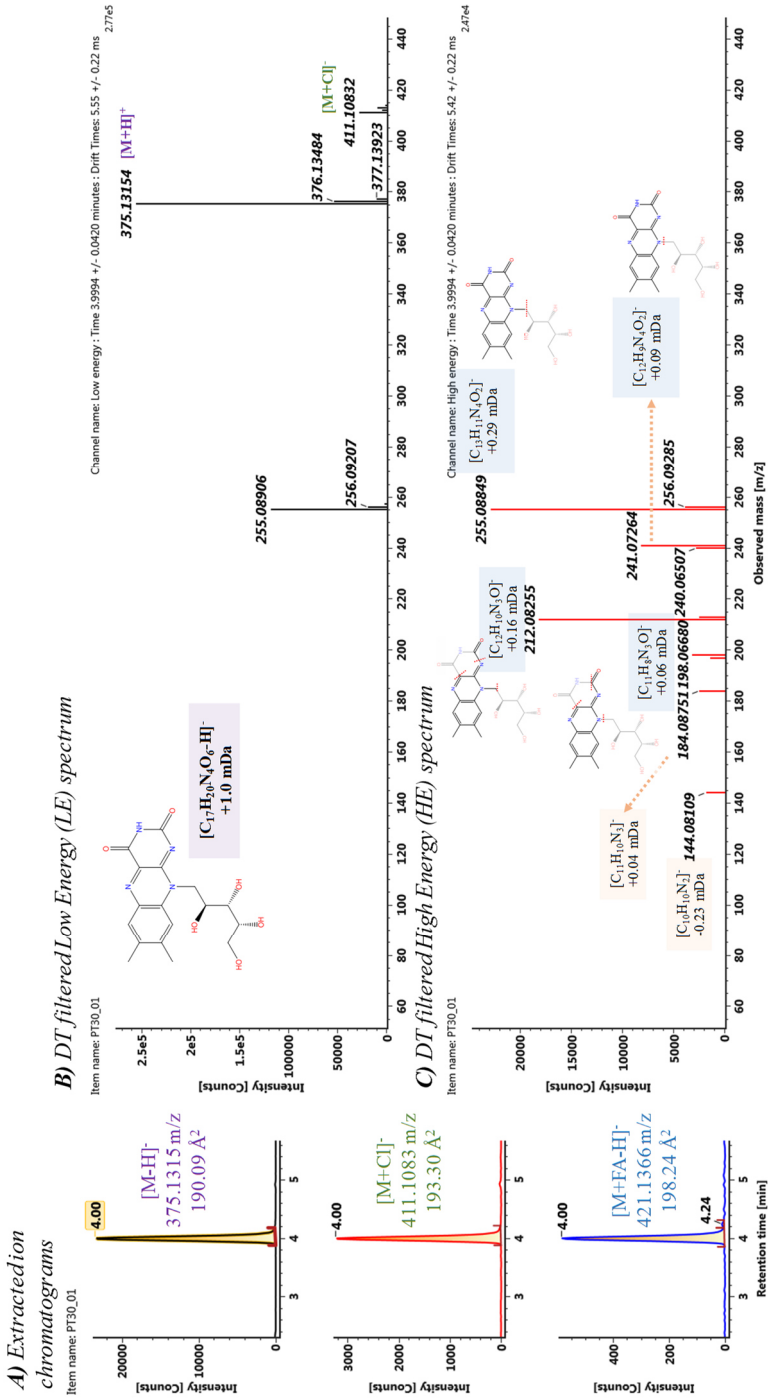


Figure. S3. *Elucidation of marker 1 Riboflavin (vitamin B2) based in chromatograms and HDMS^E spectra obtained from RP neg analysis: **A)** Extracted ion chromatograms of adducts found, along with the experimental m/z and CCS, **B)** DT filtered LE spectrum and **C)** DT filtered HE spectrum. A structure fragmentation is proposed.*

Marker 8, 10 and 11 tentatively elucidated as 4,5-didehydro-5-deoxyadenosine, adenosine-5'-(O-methylphosphate) and adenosylmethionine, respectively, shared the same fragment *m/z* 136.0619 ($[C_5H_6N_5O]^+$, -0.5 mDa) in positive ionisation mode corresponding to the adenosine base shared by all of them.

But the clearest example of similar behaviour in fragmentation occurs in the 17 tentatively elucidated lysophospholipids. This type of compounds is made up of a glycerol molecule to which a fatty acid and phosphate group are bind. Phosphate is attached via phosphodiester bond to other molecules, such as ethanolamine (LysoPE), choline (LysoPC), inositol (LysoPI) or serine (LysoPS). Due to the large number of isomers in this type of molecules (because of two position of fatty acid attachment to the glycerol molecule or number of unsaturations) it is very difficult to give a definitive identification with the data acquired for most of them. It is also difficult to obtain fragments that gives clues about the unsaturation position in the fatty acid chain, hence, the most part of its elucidation is based on the common fragmentation spectrum between the different types of lysophospholipids and the molecular formula to define the number of carbons and the unsaturations of the attached fatty acid. LysoPE can be observed in both positive and negative ionisation, and the fragmentation is characterised by neutral loss of phosphoethalonamine group (Δ*m/z* 141. 0191) in positive ionisation mode and fragments *m/z* 196.0375 and *m/z* 140.0119 in negative ionisation mode. LysoPC are typically observed in positive ionisation mode and shared fragments *m/z* 184.0740 and *m/z* 104.1075 and they correspond to phosphocholine and choline fragments. Regarding LysoPI and LysoPS, they were mainly observed in negative ionisation mode. In DT filtered HE spectrum of marker 31 LysoPI(20:4) the 4 typical ions for this class of phospholipids were observed: *m/z* 152.9949 ($[C_3H_6O_5P]^-$, -0.4 mDa) and *m/z* 241.0108 ($[C_6H_{10}O_8P]^-$, -0.5 mDa) coming from the phosphoinositol group, as well as the fragments coming from the fatty acid attached *m/z* 259.2419 ($[C_{19}H_{31}]^-$, -0.7 mDa) and *m/z* 303.2315 ($[C_{20}H_{31}O_2]^-$, -0.4 mDa). Finally, LysoPS presented the neutral loss of serine group (Δ*m/z* 87.0320). For 7 of the lysophospholipids, a match with CCS experimental database were obtained, with a delta error of +2.3 %, +1.3 %, +1.6 % for de-protonated

molecules of LysoPE(16:0), LysoPE(20:4) and LysoPE(22:6) respectively; and +0.1 % and -2.4 % for protonated molecules of both isomers of LysoPC (18:1) and both isomers of LysoPC(20:4) respectively.

In order to achieve more confidence in tentative identification, it was employed a CCS predictor model based on Multiple Adaptative Regression Splines (MARS) developed by our group (Celma et al. 2021 Analytical Chemistry submitted) with Travelling Wave Ion mobility instrument (TWIMS) (VION®, Waters). The relative errors obtained with this prediction tool were below ±4.05 % for protonated ion and ±5.86 % for deprotonated molecules for 95% of all CCS values tested. This predictor was applied from marker 1 to 17 without a match with experimental CCS (see **Table S1**).

For markers from 1 to 13 and fatty acids markers 16 and 17 without match with experimental CCS library, delta CCS errors ranging from 0.1 to 1.5 % for protonated ion and 0.1 to 1.9 % for deprotonated ion, increased the confidence of the identification. The only exception was docosahexaenoic acid with a delta error of -3.9 % for protonated molecule, while the data for deprotonated molecule was less than 1 %.

Nevertheless, the higher delta error value for phase II sulphate conjugated metabolites (-4.5 % and -7.9 % for sulphated metabolites from resveratrol (marker 14) and pterostilbene (markers 15), respectively) is due to the lack of these family of compounds during the building of the CCS prediction model. The low availability of pure analytical standards of these groups of compound makes difficult to obtain experimental CCS information to be included in the building of the prediction tool.

For phospholipids, a more specialised prediction tool was employed: LipidCCS CCS prediction model. Zhou et al. 2017 (Zhou et al., 2017). This prediction model was optimised from the CCS information obtained from a large set of lipids measured . The estimated error in the prediction for Travelling Wave Ion Mobility (TWIMS) with CCS polyAla calibrants, as VION®, is above ±5 % delta error. Nevertheless, it was employed this Lipid CCS instead of the in-house prediction model due to the lack of phospholipids experimental data in its building.

Therefore, for markers from 18 to 34 without a match with experimental CCS library, predicted CCS with delta errors ranged from 0.6 to 5.4 % for protonate ions and 1.7 to 4.8 % for deprotonated ions. Only, the LysoPS markers obtained a worst prediction with delta error from 5.1 to 8.3 % for their deprotonated ions. Moreover, due to the difficulty to find pure standards for isomeric lipids that differ in position geometry (cis/trans) to train the model, LipidCCS cannot accurately differentiated between isomers. Therefore, a more precise identity of these compounds cannot be reached with the information obtained.

All the work described so far, shows the interesting information that IMS separation offers to untargeted metabolomics studies, improving the very time-consuming identification step, still the main bottleneck in the metabolomics workflow. CCS is an additional structural descriptor highly reproducible and transversal between instruments, opposite to retention time, which depends on chromatographic separation and can be heavily influenced by sample matrix. For this reason, there are more and more compound databases that include these CCS values among the experimental data, and, in turn, predictive tools of great interest are being created and implemented. However, there is still work ahead to improve the information that IMS separation can offer, either by collecting CCS data from a great variety of compounds or the improvement of prediction models, apart from the instrumentation itself.

Nevertheless, the unequivocally confirmation of the identity (identification level 1) would only be reached by comparison with a reference standard. This level was obtained for the identification of 6 markers: marker 1 riboflavin, marker 3 cytidine, marker 4 methylnicotinamide, marker 6 xanthosine, marker 12 ADMA and marker 16 docosahexaenoic acid. For marker 2, two candidates were possible, hypoxanthine and allopurinol. The standard of hypoxanthine was purchased due to its high tentative identification level (level of 2a, match with spectral data base and CCS data base, delta error +0.01 %). Unfortunately, the analytical standard did not match in terms of retention time and hypoxanthine was discarded. Therefore, allopurinol has been given as tentative identification for this marker (level 2b) until definitive confirmation with its corresponding analytical standard.

3.4. Changes in liver metabolome

The untargeted metabolomics analysis showed three clearly differentiated metabolome profiles, one for rats fed with the high-fat and high-fructose diet (HF group), another one for rats fed with the same diet and treated with pterostilbene and another one for rats fed with the same diet and treated with resveratrol, as shown in **Figure 2**. Consequently, it can be stated that this methodology is an appropriate tool to seek biomarkers of treatment and to characterize the effects of the two phenolic compounds used in the present study. In **Figure 4**, **5** and **6**, the mean abundances of the highlighted markers in PT *vs* HF, RSV *vs* HF and PT *vs* RSV are shown, respectively. Regarding these figures, we would like to clarify that they are a visual way to observe the difference in intensity obtained for each marker between the two groups compared. Since the markers have been detected in different analysis and especially with different ionisation modes, the intensities obtained are not comparable between markers (e.g. the greater general intensity of a marker does not have to imply a higher concentration than another marker). For those markers detected in different analysis, the figure shows the normalised abundance coming from the analysis with higher statistical relevance. Supplementary **Figures S5 to S8** show in detail the variation in the intensity obtained across the groups of samples for each marker.

Resveratrol is a polyphenol, belonging to the group of stilbenes and pterostilbene, is its dimethoxy derivative. The present study shows a clear different effect of pterostilbene and resveratrol on hepatic metabolite profile, in spite of its similarity in terms of chemical structure. Firstly, the phase II sulphated metabolites of each compound were found as marker of their respective groups, and even a considerable difference could be observed for pterostilbene-4-O-sulfate between the supplementation of 15 mg/kg/d or 30 mg/kg/d of pterostilbene (**Figure S6**). When the rats fed with the high-fat high-fructose diet were treated with pterostilbene (**Figure 3**), significant changes were observed in three vitamins and derivatives (marker 1 riboflavin, marker 4 methylnicotinamide and marker 5 nicotinamide mononucleotide), four nucleotides/nucleosides (marker 2 allopurinol, marker 6 xanthosine, marker 8 4,5-didehydro-5-deoxyadenosine and marker 10 adenosine-5'-(o-methylphosfate)), three peptides/amino acids (marker 11

adenosylmethionine, marker 12 asymmetric dimethylarginine and marker 13 oxidised glutathione), and three organic acids (marker 16 docosahexaenoic acid, marker 17 docosapentaenoic acid and marker 7 indoleacrylic acid). In the case of rats treated with resveratrol (**Figure 4**), the changes affected to one vitamin-related compound (increased levels of marker 5 nicotinamide mononucleotide and decreased level of marker 9 1-(a-ribofuranosyl)-lumichrome and marker 4 methylnicotinamide), nucleotides/nucleosides ratio (higher level of marker 10 adenosine-5'-(o-methylphosfate) and lower of marker 3 cytidine), three organic acids (marker 16 docosahexaenoic acid, marker 17 docosapentaenoic acid and marker 7 indoleacrylic acid) and seventeen phospholipids. In a reduced number of cases both phenolic compounds induced similar changes in the same metabolites in comparison with HF group: decreased levels of cytidine, methylnicotinamide and indoleacrylic acid, and increased levels in nicotinamide mononucleotide, adenosine-5'-(o-methylphosphate), docosapentaenoic acid (DPA) and docosahexaenoic acid (DHA). The positive effects in different liver related diseases have been already reported for DPA and DHA (Enguita et al., 2019) as well as for nicotinamide mononucleotide (Zong et al., 2021).

Regarding the rats treated with pterostilbene, the most interesting effect was the decrease in asymmetric dimethylarginine (ADMA) and the increase in s-adenosylmethionine (SAM) in comparison to the group without supplementation (HF). ADMA, an endogenous inhibitor of nitric oxide synthase, is formed by methylation of arginine residues in proteins. These results on the positive effect of supplementation with pterostilbene are encouraging, since the reduction in ADMA are associated with adverse vascular effects associated with various clinical settings as coronary heart disease, hypertension and diabetes among others (Maas, 2005), and at the same time it is increased the presence of SAM and its possible protection against the free radical toxicity (Vergani et al., 2020).

In the case of the rats treated with resveratrol, the main changes took place in phospholipids. These molecules can be precursors of lipid mediators, which play important roles in external and internal communication and modulate cellular responses. Previous studies have demonstrated a reduction in phosphatidylcholine (PC) among other phospholipids in non-alcoholic

fatty liver disease (NAFLD) patients (Piras et al., 2021), therefore it is promising that resveratrol supplementation increases the levels of these compounds.

The present study may help to generate new hypothesis about the preventive effects of pterostilbene and resveratrol, by indicating potential pathways of interest to be addressed in further research.

4. Conclusions

The great potential of the inclusion of IMS into the conventional LC-HRMS for untargeted metabolomic studies has been demonstrated with the tentative elucidation of 34 markers related to the treatment with resveratrol and pterostilbene to rat with liver showing steatosis: 2 phase II sulphated metabolites, 17 lysophospholipids, 3 fatty acids, 4 vitamin and related compounds, 5 nucleosides/nucleotides and 3 peptide/amino acids related compounds. The implementation of CCS as additional molecular descriptor and the tools created around this value have facilitate the compound identification, going one step further in helping widen the bottleneck that the elucidation process represents for untargeted metabolomics. Significantly different results have been obtained for the two types of supplementation despite the similarity between both phenolic compounds (resveratrol and pterostilbene), although both changes are associated with beneficial health effects based on a preliminary bibliographic search of the metabolites found.

Currently, a parallel research based on the markers found in this work is being carried out in order to investigate in-depth the implication of the highlighted metabolites in stopping the progression of liver steatosis to more severe pathologies and the metabolic pathways altered.

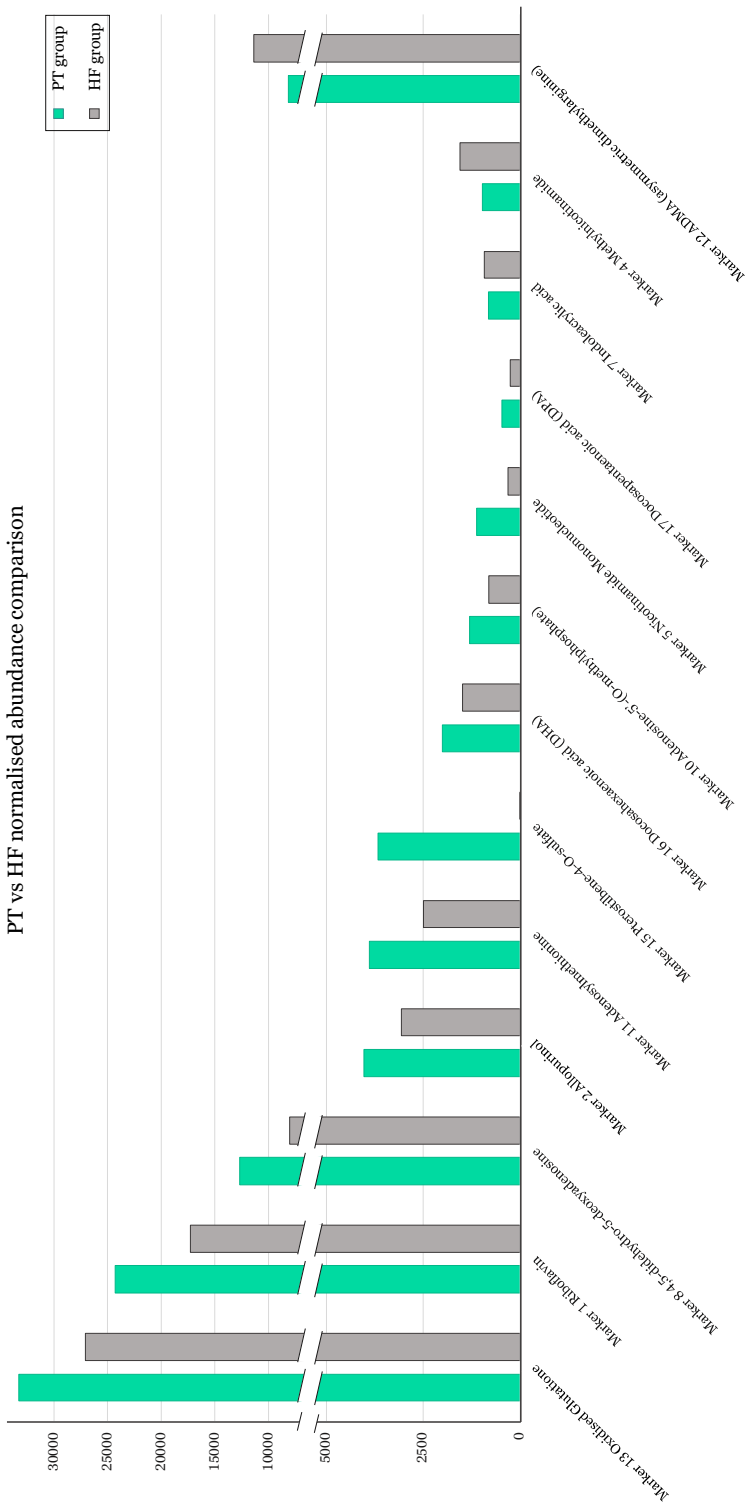


Figure. 4. *Comparison of normalised abundance means from the highlighted markers between the facing of PT group of samples (high-fat and high-fructose diet enhanced with 15 mg/kg/day or 30 mg/kg/day of pterostilbene ■) and HF group (high-fat and high-fructose diet ■). If the marker is present in more than one analysis, the normalised abundance comes from the marker higher statistical relevance.*

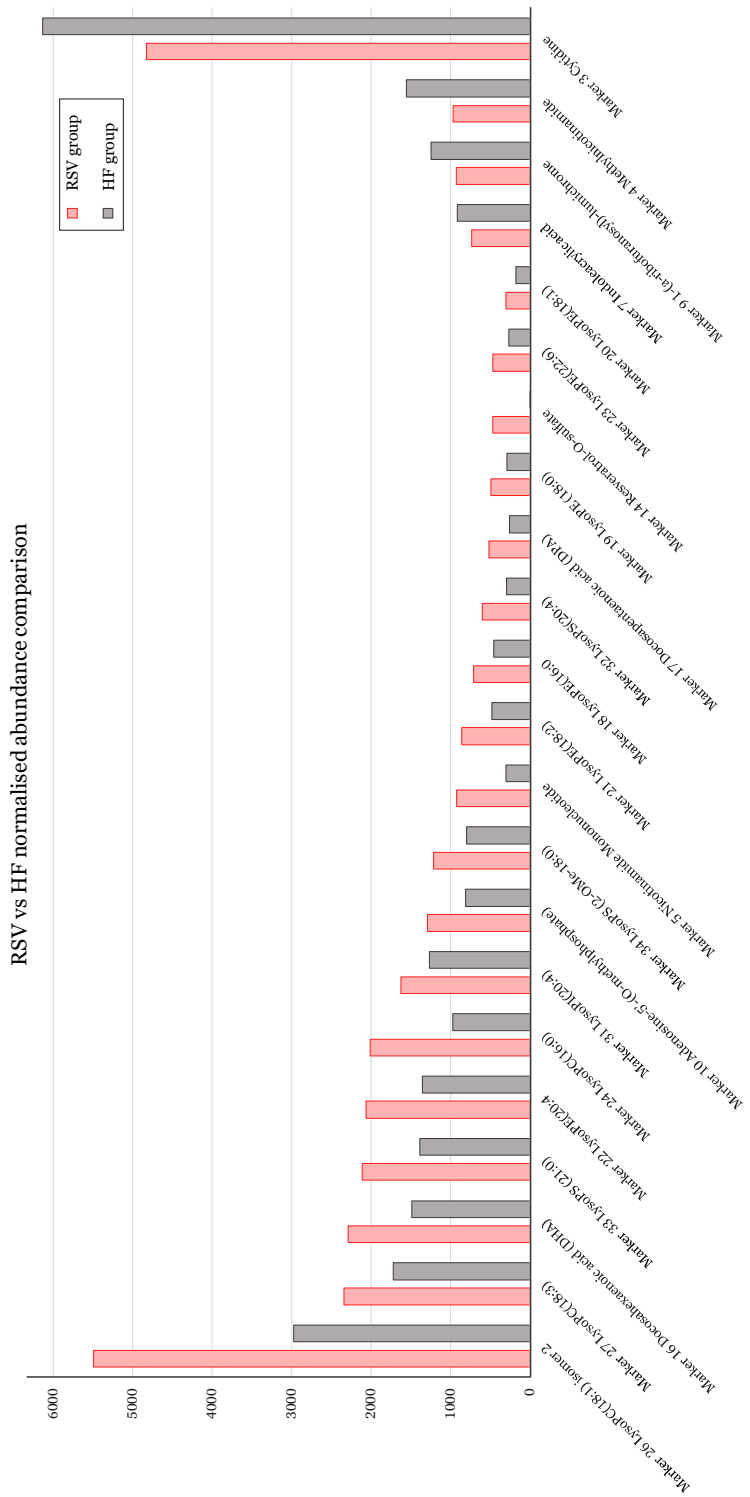


Figure. 5. *Comparison of normalised abundance means from the highlighted markers between the facing of RSV group of samples (high-fat and high-fructose diet enhanced with 30 mg/kg/day of resveratrol ■) and HF group (high-fat and high-fructose diet ■). If the marker is present in more than one analysis, the normalised abundance comes from the marker higher statistical relevance.*

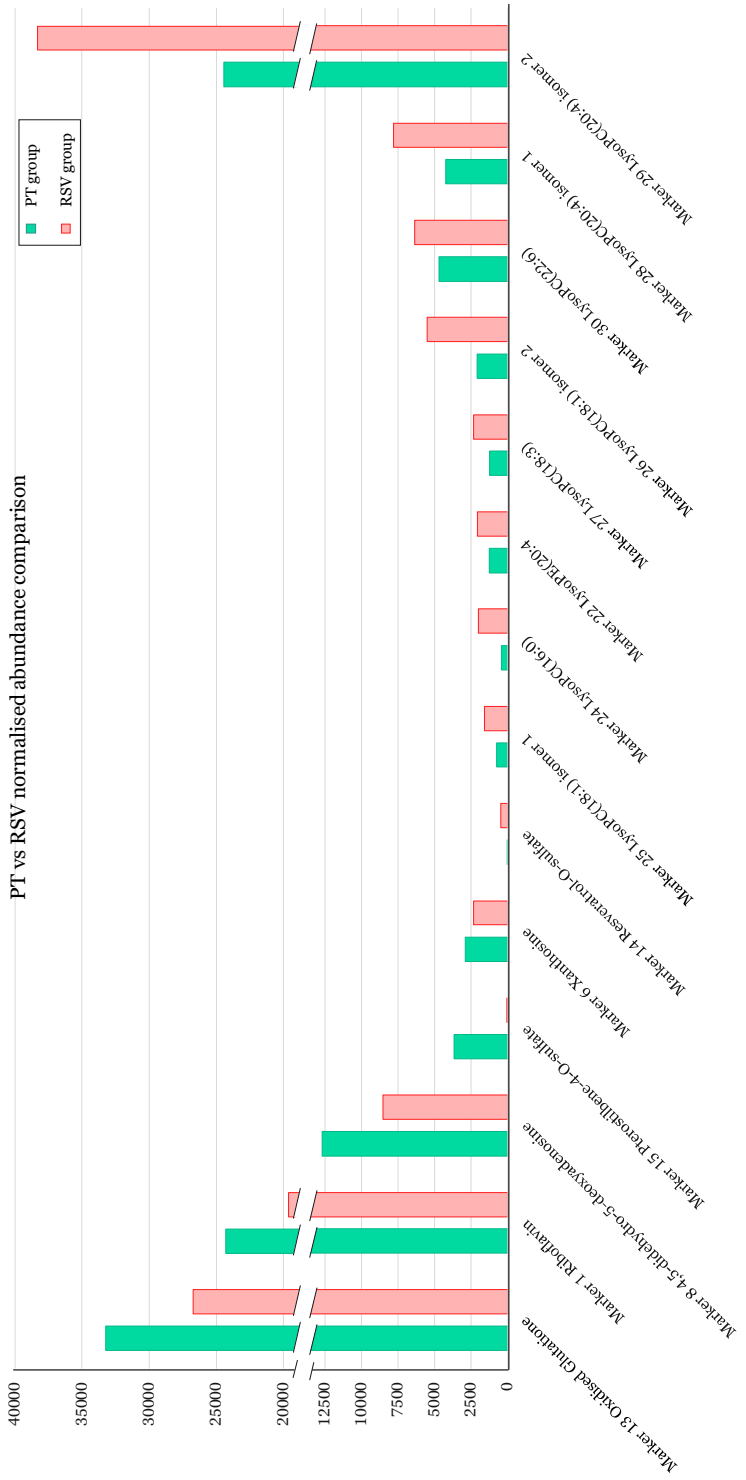


Figure. 6. *Comparison of normalised abundance means from the highlighted markers between the facing of PT group (high-fat and high-fructose diet enhanced with 15 mg/kg/day or 30 mg/kg/day of pterostilbene ■) and RSV group of samples (high-fat and high-fructose diet enhanced with 30 mg/kg/day of resveratrol ■). If the marker is present in more than one analysis, the normalised abundance comes from the marker higher statistical relevance.*

Deontological

The experiments and animal protocols were approved by the Ethical Committee of University of the Basque Country (document reference M20_2015_245 CUEID), following the European regulations (European Convention-Strasburg 1986, Directive 2003/65/EC and Recommendation 2007/526/EC).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

L Lacalle-Bergeron acknowledges the financial support of Universitat Jaume I, Spain for his pre-doctoral grant (UJI 19I001/03). D. Izquierdo-Sandoval acknowledges the Ministry of Science, Innovation and Universities of Spain for funding his research through the FPU pre-doctoral program (FPU19/ 01839). T. Portoles acknowledges Ramon y Cajal Program from the Ministry of Economy and Competitiveness, Spain (RYC-2017-22525) for funding her research. The Research Institute for Pesticides and Water (IUPA) authors acknowledge the financial support of Generalitat Valenciana, as research group of excellence PROMETEO/2019/040 and Universitat Jaume I de Castelló (UJI-B2020-25 and UJI-B2020-37). The Nutrition and Obesity Group acknowledge the financial support of the University of the Basque Country, as a research group of excellence (GIU 18/173), the *Ministerio de Economía y Competitividad-Fondo Europeo de Desarrollo Regional*, grant number AGL-2015-65719-R MINECO/FEDER, UE) and CIBERobn, grant number CB12/03/30007.

References

Bijlsma, L., Bade, R., Celma, A., Mullin, L., Cleland, G., Stead, S., Hernandez, F., & Sancho, J. V. (2017). Prediction of collision cross-section values for small molecules: application to pesticide residue analysis. Analytical Chemistry, 89(12), 6583–6589. https://doi.org/10.1021/acs.analchem.7b00741

Castro-Puyana, M., Pérez-Míguez, R., Montero, L., & Herrero, M. (2017). Application of mass spectrometry-based metabolomics approaches for food safety, quality and traceability. TrAC Trends in Analytical Chemistry, 93, 102–118. https://doi.org/10.1016/j.trac.2017.05.004

Celma, A., Bade, R., Sancho, J. V., Hernández, F., Humphries M., & Bijlsma, L. (2021) Dual prediction of retention time and collision cross section of emergin contaminants for environmental analyses. Analytical Chemistry. Submitted

Celma, A., Sancho, J. V., Schymanski, E. L., Fabregat-Safont, D., Ibáñez, M., Goshawk, J., Barknowitz, G., Hernández, F., & Bijlsma, L. (2020). Improving target and suspect screening high-resolution mass spectrometry workflows in environmental analysis by ion mobility separation. Environmental Science and Technology. https://doi.org/10.1021/acs.est.0c05713

Cifuentes, A. (2013). Foodomics: principles and applications. In A. Cifuentes (Ed.), Foodomics: advanced mass spectrometry in modern food sciences and nutrition (p. 580). John Wiley & Sons, Inc. https://www.wiley.com/en-us/Foodomics%3A+Advanced+Mass+Spectrometry+in+Modern+Food+Science+and+Nutrition-p-9781118169452

Enguita, M., Razquin, N., Pamplona, R., Quiroga, J., Prieto, J., & Fortes, P. (2019). The cirrhotic liver is depleted of docosahexaenoic acid (DHA), a key modulator of NF-κB and TGFβ pathways in hepatic stellate cells. Cell Death & Disease, 10(1), 14. https://doi.org/10.1038/s41419-018-1243-0

Fu, J., Zhang, L.-L., Li, W., Zhang, Y., Zhang, Y., Liu, F., & Zou, L. (2022). Application of metabolomics for revealing the interventional effects of functional foods on metabolic diseases. Food Chemistry, 367(April 2021), 130697. https://doi.org/10.1016/j.foodchem.2021.130697

Gimeno-Mallench, L., Mas-Bargues, C., Inglés, M., Olaso, G., Borras, C., Gambini, J., & Vina, J. (2019). Resveratrol shifts energy metabolism to increase lipid oxidation in healthy old mice. Biomedicine &

Pharmacotherapy, 118(March), 109130. https://doi.org/10.1016/j.biopha.2019.109130

Gómez-Zorita, S., González-Arceo, M., Trepiana, J., Aguirre, L., Crujeiras, A. B., Irles, E., Segues, N., Bujanda, L., & Portillo, M. P. (2020). Comparative effects of pterostilbene and its parent compound resveratrol on oxidative stress and inflammation in steatohepatitis induced by high-fat high-fructose feeding. Antioxidants, 9(11), 1042. https://doi.org/10.3390/antiox9111042

Kapetanovic, I. M., Muzzio, M., Huang, Z., Thompson, T. N., & McCormick, D. L. (2011). Pharmacokinetics, oral bioavailability, and metabolic profile of resveratrol and its dimethylether analog, pterostilbene, in rats. Cancer Chemotherapy and Pharmacology, 68(3), 593–601. https://doi.org/10.1007/s00280-010-1525-4

Karu, N., Deng, L., Slae, M., Guo, A. C., Sajed, T., Huynh, H., Wine, E., & Wishart, D. S. (2018). A review on human fecal metabolomics: Methods, applications and the human fecal metabolome database. Analytica Chimica Acta, 1030, 1–24. https://doi.org/10.1016/j.aca.2018.05.031

Koh, Y.-C., Ho, C.-T., & Pan, M.-H. (2021). Recent Advances in Health Benefits of Stilbenoids. Journal of Agricultural and Food Chemistry, 69(35), 10036–10057. https://doi.org/10.1021/acs.jafc.1c03699

Lacalle-Bergeron, L., Izquierdo-Sandoval, D., Sancho, J. V., López, F. J., Hernández, F., & Portolés, T. (2021). Chromatography hyphenated to high resolution mass spectrometry in untargeted metabolomics for investigation of food (bio)markers. TrAC Trends in Analytical Chemistry, 135, 116161. https://doi.org/10.1016/j.trac.2020.116161

Lacalle-Bergeron, L., Portolés, T., López, F. J., Sancho, J. V., Ortega-Azorín, C., Asensio, E. M., Coltell, O., & Corella, D. (2020). Ultra-Performance Liquid Chromatography-Ion Mobility Separation-Quadruple Time-of-Flight MS (UHPLC-IMS-QTOF MS) Metabolomics for Short-Term Biomarker Discovery of Orange Intake: A Randomized, Controlled Crossover Study. Nutrients, 12(7), 1916. https://doi.org/10.3390/nu12071916

Lewinska, M., Santos-Laso, A., Arretxe, E., Alonso, C., Zhuravleva, E., Jimenez-Agüero, R., Eizaguirre, E., Pareja, M. J., Romero-Gómez, M., Arrese, M., Suppli, M. P., Knop, F. K., Oversoe, S. K., Villadsen, G. E., Decaens, T., Carrilho, F. J., de Oliveira, C. P., Sangro, B., Macias, R. I. R., … Andersen, J. B. (2021). The altered serum lipidome and its diagnostic potential for Non-Alcoholic Fatty Liver (NAFL)-associated hepatocellular carcinoma. EBioMedicine, 73, 103661. https://doi.org/10.1016/j.ebiom.2021.103661

Maas, R. (2005). Pharmacotherapies and their influence on asymmetric dimethylargine (ADMA). Vascular Medicine, 10(2_suppl), S49–S57. https://doi.org/10.1191/1358863x05vm605oa

Madatali Abuwani, A., Priyadarshini Dash, S., Ganesan, R., Renu, K., Vellingiri, B., Kandasamy, S., C.R., S. R., & Valsala Gopalakrishnan, A. (2021). Gut microbiome and metabolic response in non-alcoholic fatty liver disease. Clinica Chimica Acta, 523(October), 304–314. https://doi.org/10.1016/j.cca.2021.10.014

Mairinger, T., Causon, T. J., & Hann, S. (2018). The potential of ion mobility–mass spectrometry for non-targeted metabolomics. In Current Opinion in Chemical Biology (Vol. 42, pp. 9–15). Elsevier Ltd. https://doi.org/10.1016/j.cbpa.2017.10.015

Paglia, G., & Astarita, G. (2017). Metabolomics and lipidomics using traveling-wave ion mobility mass spectrometry. Nature Protocols, 12(4), 797–813. https://doi.org/10.1038/nprot.2017.013

Piras, C., Noto, A., Ibba, L., Deidda, M., Fanos, V., Muntoni, S., Leoni, V. P., & Atzori, L. (2021). Contribution of Metabolomics to the Understanding of NAFLD and NASH Syndromes: A Systematic Review. Metabolites, 11(10), 694. https://doi.org/10.3390/metabo11100694

Pirola, C. J., & Sookoian, S. (2018). Multiomics biomarkers for the prediction of nonalcoholic fatty liver disease severity. World Journal of Gastroenterology, 24(15), 1601–1615. https://doi.org/10.3748/wjg.v24.i15.1601

Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H. P., & Hollender, J. (2014). Identifying Small Molecules via High Resolution

Mass Spectrometry: Communicating Confidence. Environmental Science & Technology, 48(4), 2097–2098. https://doi.org/10.1021/es5002105

Segers, K., Declerck, S., Mangelings, D., Heyden, Y. Vander, & Eeckhaut, A. Van. (2019). Analytical techniques for metabolomic studies: A review. Bioanalysis, 11(24), 2297–2318. https://doi.org/10.4155/bio-2019-0014

Vergani, L., Baldini, F., Khalil, M., Voci, A., Putignano, P., & Miraglia, N. (2020). New perspectives of S-adenosylmethionine (SAMe) applications to attenuate fatty acid-induced steatosis and oxidative stress in hepatic and endothelial Cells. Molecules, 25(18), 4237. https://doi.org/10.3390/molecules25184237

Vivanco, F., Barderas, M. G., Laborde, C. M., Posada, M., De La Cuesta, F., Zubiri, I., & Alvarez-Llamas, G. (2011). Metabolomic profiling for identification of novel potential biomarkers in cardiovascular diseases. In Journal of Biomedicine and Biotechnology (Vol. 2011, pp. 1–9). https://doi.org/10.1155/2011/790132

Wang, T. J., Larson, M. G., Vasan, R. S., Cheng, S., Rhee, E. P., McCabe, E., Lewis, G. D., Fox, C. S., Jacques, P. F., Fernandez, C., O'Donnell, C. J., Carr, S. A., Mootha, V. K., Florez, J. C., Souza, A., Melander, O., Clish, C. B., & Gerszten, R. E. (2011). Metabolite profiles and the risk of developing diabetes. Nature Medicine, 17(4), 448–453. https://doi.org/10.1038/nm.2307

Wolfender, J.-L., Marti, G., Thomas, A., & Bertrand, S. (2015). Current approaches and challenges for the metabolite profiling of complex natural extracts. Journal of Chromatography A, 1382, 136–164. https://doi.org/10.1016/j.chroma.2014.10.091

Worley, B., & Powers, R. (2012). Multivariate Analysis in Metabolomics. Current Metabolomics, 1(1), 92–107. https://doi.org/10.2174/2213235X130108

Zhou, Z., Tu, J., Xiong, X., Shen, X., & Zhu, Z.-J. (2017). LipidCCS: prediction of collision cross-section values for lipids with high precision to support ion mobility–mass spectrometry-based lipidomics. Analytical

Chemistry, 89(17), 9559–9566. https://doi.org/10.1021/acs.analchem.7b02625

Zong, Z., Liu, J., Wang, N., Yang, C., Wang, Q., Zhang, W., Chen, Y., Liu, X., & Deng, H. (2021). Nicotinamide mononucleotide inhibits hepatic stellate cell activation to prevent liver fibrosis via promoting PGE2 degradation. Free Radical Biology and Medicine, 162(November 2020), 571–581. https://doi.org/10.1016/j.freeradbiomed.2020.11.014

Table S1. *Experimental data for the 34 markers tentatively identified.*

Nº	Identification level [a]	Compound	Elemental composition	Feature name	Rt (min)	de/protonated molecule *m/z*	Mass error (mDa/ppm)	Experimental CCS (Å²) de/protonated molecule *m/z*	CCS delta error (%) de/protonated molecule *m/z*	Adducts detected
1	1	Riboflavin (Vitamin B2)	$C_{17}H_{20}N_4O_6$	HILICpos_2.05_376.1390n	2.05	377.1463	+0.2/+0.5	186.02	+0.2 %[b]	$M+H]^+$ $[M-H_2O+H]^+$ $[M+Na]^+$ $[M+K]^+$
				RPpos_4.00_376.1388n	4.00	377.1461	-0.01/-0.02	185.82	+0.1 %[b]	$[M+H]^+$ $[M-H_2O+H]^+$ $[M+Na]^+$ $[M+K]^+$
				RPneg_3.99_376.1386n	3.99	375.1313	+0.8/+2.2	190.09	+1.4 %[c]	$[M-H]^-$ $[M-Cl]^-$ $[M-FA-H]^-$
2	2b	Allopurinol	$C_5H_4N_4O$	HILICpos_2.36_137.0460m/z	2.36	137.0460	-0.3/-2.5	123.13	+0.1 %[c]	$[M+H]^+$
3	1	Cytidine	$C_9H_{13}N_3O_5$	HILICpos_3.18_243.0860n	3.18	244.0933	-0.1/-0.2	223.12	+0.4 %[b]	$[M+H]^+$ $[M+Na]^+$ $[M+K]^+$ $[2M+H]^+$ $[2M+Na]^+$
4	1	Methylnicotinamide	$C_7H_8N_2O$	HILICpos_7.35_137.0709m/z	7.35	137.0709	-0.6/-4.3	124.45	+1.4 %[c]	$[M+H]^+$
5	2a	Nicotinamide Mononucleotide	$C_{11}H_{15}N_2O_8P$	HILICpos_7.87_335.0644m/z	7.87	335.0644	-0.0/-0.1	163.66	-0.2 %[b]	$[M+H]^+$
6	1	Xanthosine	$C_{10}H_{12}N_4O_6$	RPneg_1.13_284.0760n	1.13	283.0687	+0.8/+2.9	156.34	+1.3 % [b]	$[M-H]^-$ $[2M-H]^-$
7	2b	Indoleacrylic acid	$C_{11}H_9NO_2$	RPpos_1.85_188.0707m/z	1.85	188.0707	-0.5/-2.4	140.46	-0.1 %[c]	$[M+H]^+$
8	3	4,5-didehydro-5-deoxyadenosine	$C_{10}H_{11}N_5O_3$	HILICpos_1.25_250.0940m/z	1.25	250.0940	-0.0/-0.1	153.61	+1.5 %[c]	$[M+H]^+$
9	3	1-(a-ribofuranosyl)-lumichrome	$C_{17}H_{18}N_4O_6$	HILICpos_1.28_375.1304m/z	1.28	375.1304	+0.1/+0.16	183.61	+0.5 %[c]	$[M+H]^+$

[a] *Identification level according to Celma et al. (2020)*

[b] *CCS delta error (%) by comparison with the experimental CCS from Metabolic Profiling CCS library for Progenesis QI (measured in Waters TWIMS instruments)*

[c] *CCS delta error (%) by comparison with the predicted CCS obtained by CCS prediction tool MARS-based developed by Celma et al. (2021), lipids and sulfates metabolites error in prediction not published yet*

[d] *CCS delta error (%) by comparison with the predicted CCS for protonated molecule obtained by LipidCCS prediction tool developed by Zhou et al. (2017)*

Table S1. *Experimental data for the 34 markers tentatively identified (continuation).*

Nº	Identification level [a]	Compound	Elemental composition	Feature name	Rt (min)	de/protonated molecule *m/z*	Mass error (mDa/ppm)	Experimental CCS (Å²) de/protonated molecule *m/z*	CCS delta error (%) de/protonated molecule *m/z*	Adducts detected
10	3	Adenosine-5'-(O-methylphosphate)	$C_{11}H_{16}N_5O_7P$	HILICpos_4.78_361.0794n	4.78	362.0867	+0.1/+0.4	169.76	-3.2 %[c]	$[M+H]^+$ $[M+Na]^+$
				HILICneg_4.77_360.0718m/z	4.77	360.0718	+0.9/+2.5	178.57	-1.9 %[c]	$[M-H]^-$
				RPpos_0.97_362.0866m/z	0.97	362.0866	+0.0/+0.1	170.95	-2.5 %[c]	$[M+H]^+$
				RPneg_0.98_361.0792n	0.98	360.0719	+0.9/+2.7	178.65	-1.9 %[c]	$[M-H]^-$ $[2M-H]^-$
11	2b	Adenosylmethionine	$C_{15}H_{22}N_6O_5S$	HILICpos_9.27_399.1448m/z	9.27	399.1448	-0.3/-0.7	185.20	-3.9 %[c]	$[M+H]^+$
12	1	ADMA (asymmetric dimethylarginine)	$C_8H_{18}N_4O_2$	HILICpos_9.15_203.1505m/z	9.15	203.1505	-0.3/-1.5	143.49	+1.7 %[b]	$[M+H]^+$
13	2a	Oxidized Glutatione	$C_{20}H_{32}N_6O_{12}S_2$	HILICpos_6.54_612.1526n	6.54	613.1599	+0.1/+0.2	223.81	-1.4 %[b]	$[M+H]^+$ $[M+Na]^+$ $[M+K]^+$ $[M+2H]^+$
				HILICneg_6.62_611.1448m/z	6.62	611.1448	+0.7/+1.0	219.25	+0.8 %[b]	$[M-H]^-$
				RPpos_0.75_612.1528n	0.75	613.1601	+0.3/+0.5	223.80	-1.5 %[b]	$[M+H]^+$ $[M+Na]^+$ $[M+K]^+$ $[M+2H]^+$
				RPneg_0.78_611.1452m/z	0.78	611.1452	+1.0/+1.7	219.74	+0.8 %[b]	$[M-H]^-$
14	3	Resveratrol-O-sulfate	$C_{14}H_{12}O_6S$	HILICneg_0.55_307.0281m/z	0.55	307.0281	+0.5/+1.5	177.13	-4.5 %[c]	$[M-H]^-$
15	3	Pterostilbene-4'-O-sulfate	$C_{16}H_{16}O_6S$	HILICneg_0.55_335.0597m/z	0.55	335.0597	+0.8/+2.3	193.05	-7.9 %[c]	$[M-H]^-$

[a] *Identification level according to Celma et al. (2020)*

[b] *CCS delta error (%) by comparison with the experimental CCS from Metabolic Profiling CCS library for Progenesis QI (measured in Waters TWIMS instruments)*

[c] *CCS delta error (%) by comparison with the predicted CCS obtained by CCS prediction tool MARS-based developed by Celma et al. (2021), lipids and sulfates metabolites error in prediction not published yet*

[d] *CCS delta error (%) by comparison with the predicted CCS for protonated molecule obtained by LipidCCS prediction tool developed by Zhou et al. (2017)*

Table S1. *Experimental data for the 34 markers tentatively identified (continuation).*

Nº	Identification level [a]	Compound	Elemental composition	Feature name	Rt (min)	de/protonated molecule *m/z*	Mass error (mDa/ppm)	Experimental CCS (Å²) de/protonated molecule *m/z*	CCS delta error (%) de/protonated molecule *m/z*	Adducts detected
16	1	Docosahexaenoic acid (DHA)	$C_{22}H_{32}O_2$	HILICneg_0.76_327.2331m/z	0.76	327.2331	+0.7/+2.1	199.47	+1.9 %[c]	$[M-H]^-$
				RPpos_15.30_328.2403n	15.30	329.2476	-0.5/-1.4	187.95	-4.3 %[c]	$[M+H]^+$ $[M+Na]^+$ $[M+K]^+$
				RPneg_15.32_327.2332m/z	15.32	327.2332	+0.8/+2.4	197.16	+0.8 %[c]	$[M-H]^-$
17	2a	Docosapentaenoic acid (DPA)	$C_{22}H_{34}O_2$	HILICneg_0.76_329.2483m/z	0.76	329.2483	+0.2/+0.7	199.14	-0.1 %[c]	$[M-H]^-$
				RPneg_15.64_329.2488m/z	15.64	329.2488	+0.7/+2.2	196.75	+0.4 %[c]	$[M-H]^-$
18	3	LysoPE(16:0)	$C_{21}H_{44}NO_7P$	RPneg _14.60_452.2785m/z	14.60	452.2785	+0.8/+1.7	217.53	+2.3 %[b]	$[M-H]^-$
19	3	LysoPE (18:0)	$C_{23}H_{48}NO_7P$	RPneg_14.56_480.3098m/z	14.56	480.3098	+0.8/+1.6	225.20	+4.8 %[d]	$[M-H]^-$
20	3	LysoPE(18:1)	$C_{23}H_{46}NO_7P$	HILICpos_4.13_480.3079m/z	4.13	480.3079	-1.1/-2.3	222.07	+1.4 %[d]	$[M+H]^+$
				HILICneg_4.16_478.2929m/z	4.16	478.2929	-0.5/-1.0	221.62	+3.9 %[d]	$[M-H]^-$
21	3	LysoPE(18:2)	$C_{23}H_{44}NO_7P$	HILICneg_4.19_476.2783m/z	4.19	476.2783	+0.6/+1.2	219.53	+3.5 %[d]	$[M-H]^-$
22	2a	LysoPE(20:4)	$C_{25}H_{44}NO_7P$	HILICpos_4.09_501.2860n	4.09	502.2933	-0.1/-0.1	217.29	+1.8 %[d]	$[M+H]^+$ $[M+Na]^+$ $[M+K]^+$
				HILICneg_4.15_500.2784m/z	4.15	500.2784	+0.7/+1.4	221.85	+1.3 %[b]	$[M-H]^-$
				RPneg_14.27_500.2786m/z	14.27	500.2786	+0.9/+1.8	222.08	+1.4 %[b]	$[M-H]^-$

[a] *Identification level according to Celma et al. (2020)*

[b] *CCS delta error (%) by comparison with the experimental CCS from Metabolic Profiling CCS library for Progenesis QI (measured in Waters TWIMS instruments)*

[c] *CCS delta error (%) by comparison with the predicted CCS obtained by CCS prediction tool MARS-based developed by Celma et al. (2021), lipids and sulfates metabolites error in prediction not published yet*

[d] *CCS delta error (%) by comparison with the predicted CCS for protonated molecule obtained by LipidCCS prediction tool developed by Zhou et al. (2017)*

Table S1. *Experimental data for the 34 markers tentatively identified (continuation).*

Nº	Identification level [a]	Compound	Elemental composition	Feature name	Rt (min)	de/protonated molecule *m/z*	Mass error (mDa/ppm)	Experimental CCS (Å²) de/protonated molecule *m/z*	CCS delta error (%) de/protonated molecule *m/z*	Adducts detected
23	2a	LysoPE(22:6)	$C_{27}H_{44}NO_7P$	HILICpos_4.08_525.2857n	4.08	526.2930	-0.4/-0.7	220.50	+0.6 %[d]	$[M+H]^+$ $[M+Na]^+$ $[M+K]^+$
				HILICneg_4.13_524.2783m/z	4.13	524.2783	+0.6/+1.1	226.51	+1.6 %[b]	$[M-H]^-$
24	3	LysoPC(16:0)	$C_{24}H_{50}NO_7P$	HILICpos_4.80_495.3325n	4.80	496.3398	-0.5/-1.0	237.00	+1.9 %[d]	$[M+H]^+$ $[M+Na]^+$
				RPpos_14.57_495.3327n	14.57	496.3400	-0.3/-0.6	236.34	+1.7%[d]	$[M+H]^+$ $[M+Na]^+$ $[M+K]^+$
25	2a	LysoPC(18:1) isomer 1	$C_{26}H_{52}NO_7P$	HILICpos_4.56_522.3535m/z	4.65	522.3535	-2.5/-4.7	241.09	+0.1 %[b]	$[M+H]^+$
26	3	LysoPC(18:1) isomer 2	$C_{26}H_{52}NO_7P$	HILICpos_4.73_522.3536m/z	4.73	522.3536	-2.4/-4.5	240.03	+0.1 %[b]	$[M+H]^+$
27	3	LysoPC(18:3)	$C_{26}H_{48}NO_7P$	HILICpos_4.80_517.3155n	4.80	518.3228	-0.5/-1.0	240.16	+5.4 %[d]	$[M+H]^+$ $[M+Na]^+$
28	2a	LysoPC(20:4) isomer 1	$C_{28}H_{50}NO_7P$	HILICpos_4.53_543.3329n	4.53	544.3402	-0.1/-0.2	236.94	-2.4 %[b]	$[M+H]^+$ $[M+Na]^+$ $[M+K]^+$
29	3	LysoPC(20:4) isomer 2	$C_{28}H_{50}NO_7P$	HILICpos_4.68_543.3325n	4.68	544.3398	-0.5/-0.9	234.85	-2.4 %[b]	$[M+H]^+$ $[M+Na]^+$ $[M+K]^+$
30	3	LysoPC(22:6)	$C_{30}H_{50}NO_7P$	HILICpos_4.65_567.3321n	4.65	568.3394	-0.9/-1.6	238.81	+2.4 %[d]	$[M+H]^+$ $[M+Na]^+$
31	3	LysoPI(20:4)	$C_{29}H_{49}O_{12}P$	HILICneg_3.92_619.2889m/z	3.92	619.2889	+0.6/+0.9	246.37	+3.1 %[d]	$[M-H]^-$
32	3	LysoPS(20:4)	$C_{26}H_{44}NO_9P$	HILICneg_4.32_544.2682m/z	4.32	544.268	+0.5/+0.8	234.25	+5.1 %[d]	$[M-H]^-$
33	3	LysoPS (21:0)	$C_{27}H_{54}NO_9P$	RPneg_14.77_566.3465m/z	14.77	566.3465	+0.7/+1.3	250.91	+8.2 %[d]	$[M-H]^-$
34	3	LysoPS (2-OMe-18:0)	$C_{25}H_{52}NO_9P$	RPneg_14.56_540.3310m/z	14.56	540.331	+0.9/+1.6	245.49	+8.3 %[d]	$[M-H]^-$

[a] *Identification level according to Celma et al. (2020)*

[b] *CCS delta error (%) by comparison with the experimental CCS from Metabolic Profiling CCS library for Progenesis QI (measured in Waters TWIMS instruments)*

[c] *CCS delta error (%) by comparison with the predicted CCS obtained by CCS prediction tool MARS-based developed by Celma et al. (2021), lipids and sulfates metabolites error in prediction not published yet*

[d] *CCS delta error (%) by comparison with the predicted CCS for protonated molecule obtained by LipidCCS prediction tool developed by Zhou et al. (2017)*

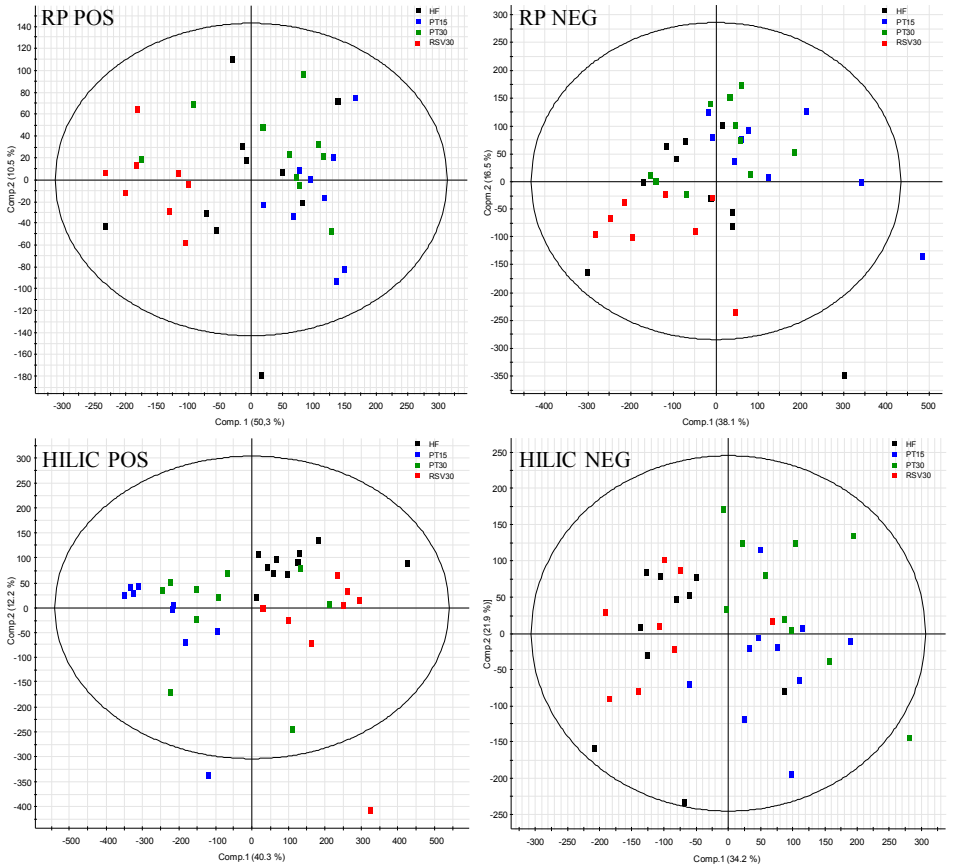


Figure. S1. *PCA score plot component 1 vs component 2 for liver metabolic profiles in the four different analysis modes. Four groups of liver samples from Wistar rats fed with different diets were analysed: high-fat and high-fructose diet (HF ■), and high-fat and high-fructose diet enhanced with 15 mg/kg/day or 30 mg/kg/day of pterostilbene (PT15 ■ and PT30 ■) or 30 mg/kg/day of resveratrol (RSV30 ■). The variance explained 80 %, 85 %, 70 % and 79 % for RP pos, RP neg, HILIC pos and HILIC neg, respectively.*

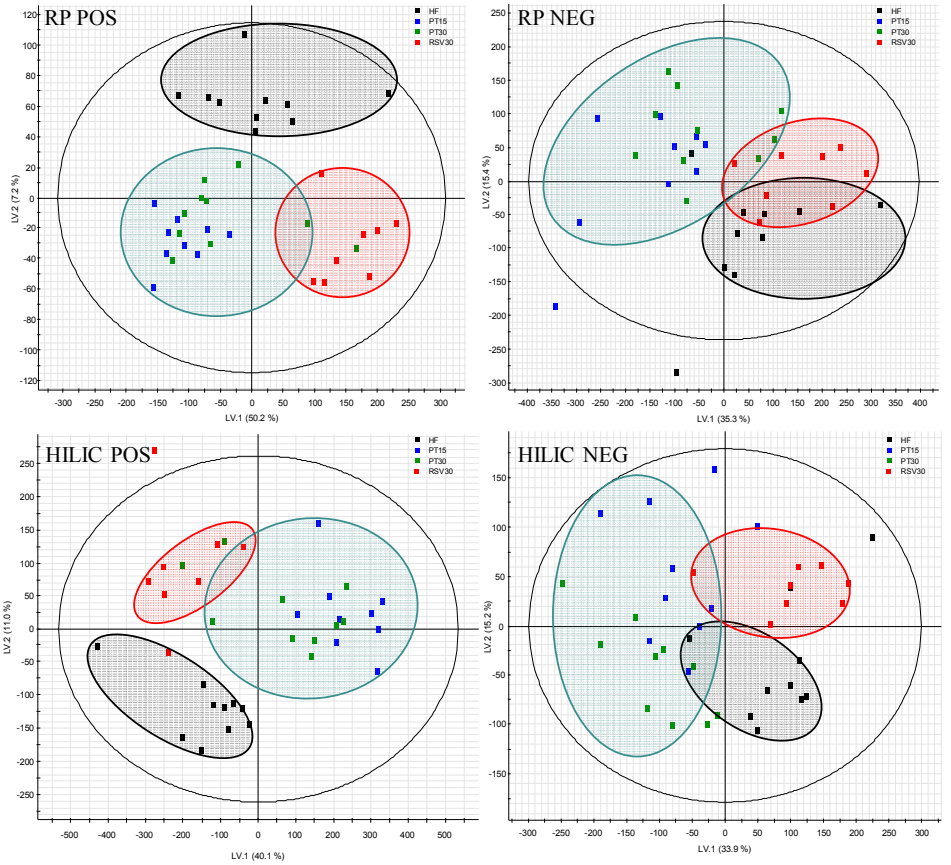


Figure. S2. *PLS-DA score plot latent variable 1 vs latent variable 2 for liver metabolic profiles in the four different analysis modes. Four groups of liver samples from Wistar rats fed with different diets were analysed: high-fat and high-fructose diet (HF ■), and high-fat and high-fructose diet enhanced with 15 mg/kg/day or 30 mg/kg/day of pterostilbene (PT15 ■ and PT30 ■) or 30 mg/kg/day of resveratrol (RSV30 ■). Being R^2 is coefficient for variance explained and Q^2 is coefficient for variance predicted, the following results were obtained: R^2Y= 48 % and Q^2 = 40 % for RP pos, R^2Y= 51 % and Q^2 = 33 % for RP neg, R^2Y= 45 % and Q^2 = 37 % for HILIC pos and R^2Y= 36 % and Q^2 = 18 % for HILIC neg.*

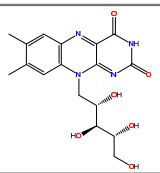
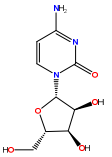
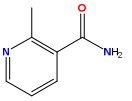
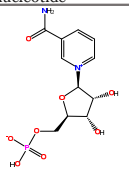
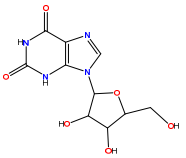
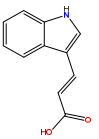
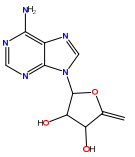
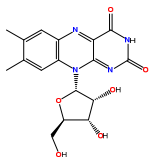
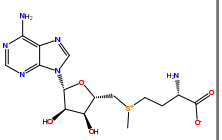
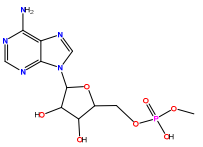
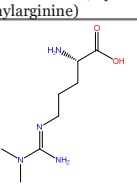
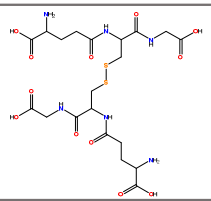
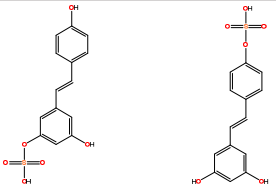
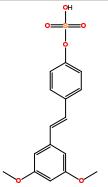
Marker 1. Riboflavin (vitamin B2)	Marker 2. Allopurinol	Marker 3. Cytidine	Marker 4. Methylnicotinamide
			
Marker 5. Nicotinamide mononucleotide	**Marker 6. Xanthosine**	**Marker 7. Indoleacrylic acid**	**Marker 8. 4,5-didehydro-5-deoxyadenosine**
			
Marker 9. 1-(a-ribofuranosyl)-lumichrome	**Marker 10. Adenosine-5'-(O-methylphosphate)**	**Marker 11. Adenosylmethionine**	**Marker 12. ADMA (asymmetric dimethylarginine)**
			
Marker 13. Oxidised Glutatione	**Marker 14. Resveratrol-O-sulfate**		**Marker 15. Pterostilbene-4-O-sulfate**
			

Figure. S4. *Tentative structures and identity for marker from 1 to 15.*

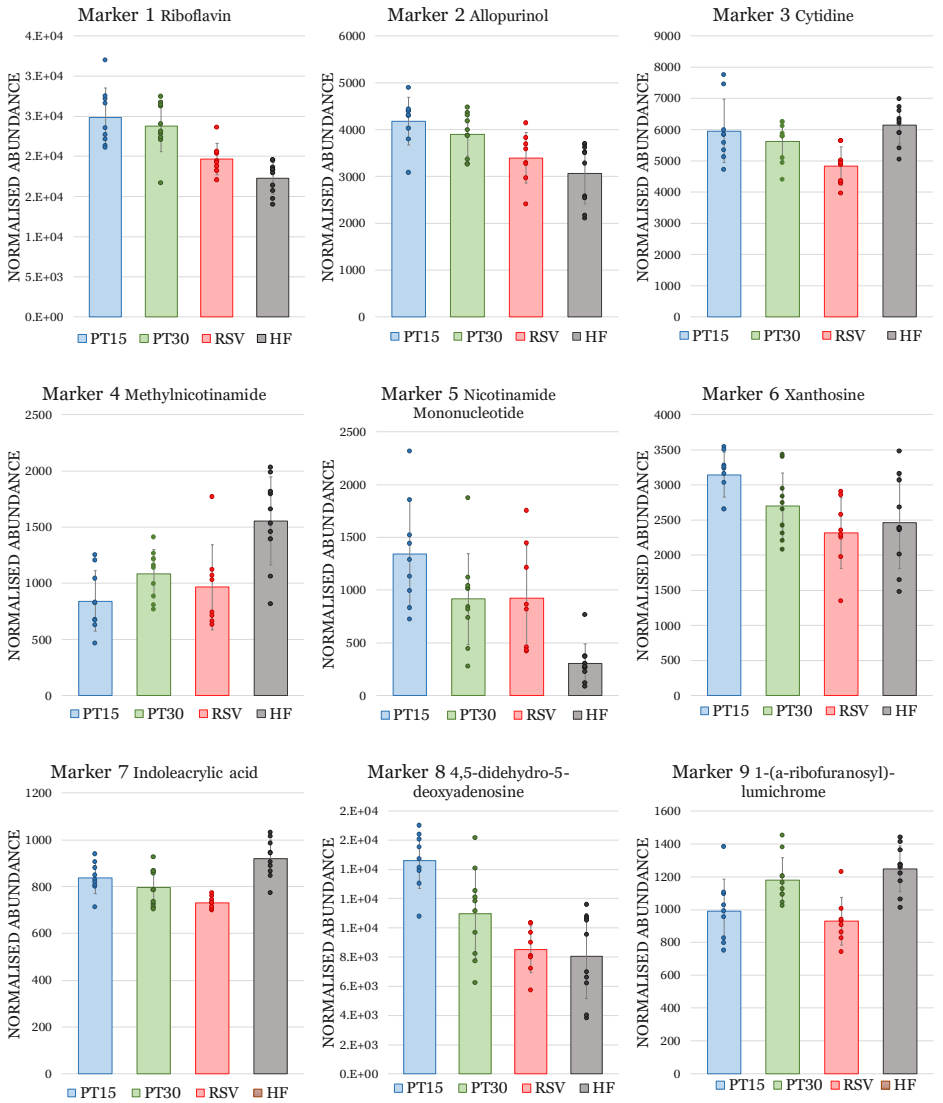


Figure S5. *Comparison of mean of normalised abundance between the different groups samples of markers from 1 to 9: high-fat and high-fructose diet (HF ■), and high-fat and high-fructose diet enhanced with 15 mg/kg/day or 30 mg/kg/day of pterostilbene (PT15 ■ and PT30 ■) or 30 mg/kg/day of resveratrol (RSV30 ■). If the marker is present in more than one analysis, the normalised abundance comes from the marker higher statistical relevance.*

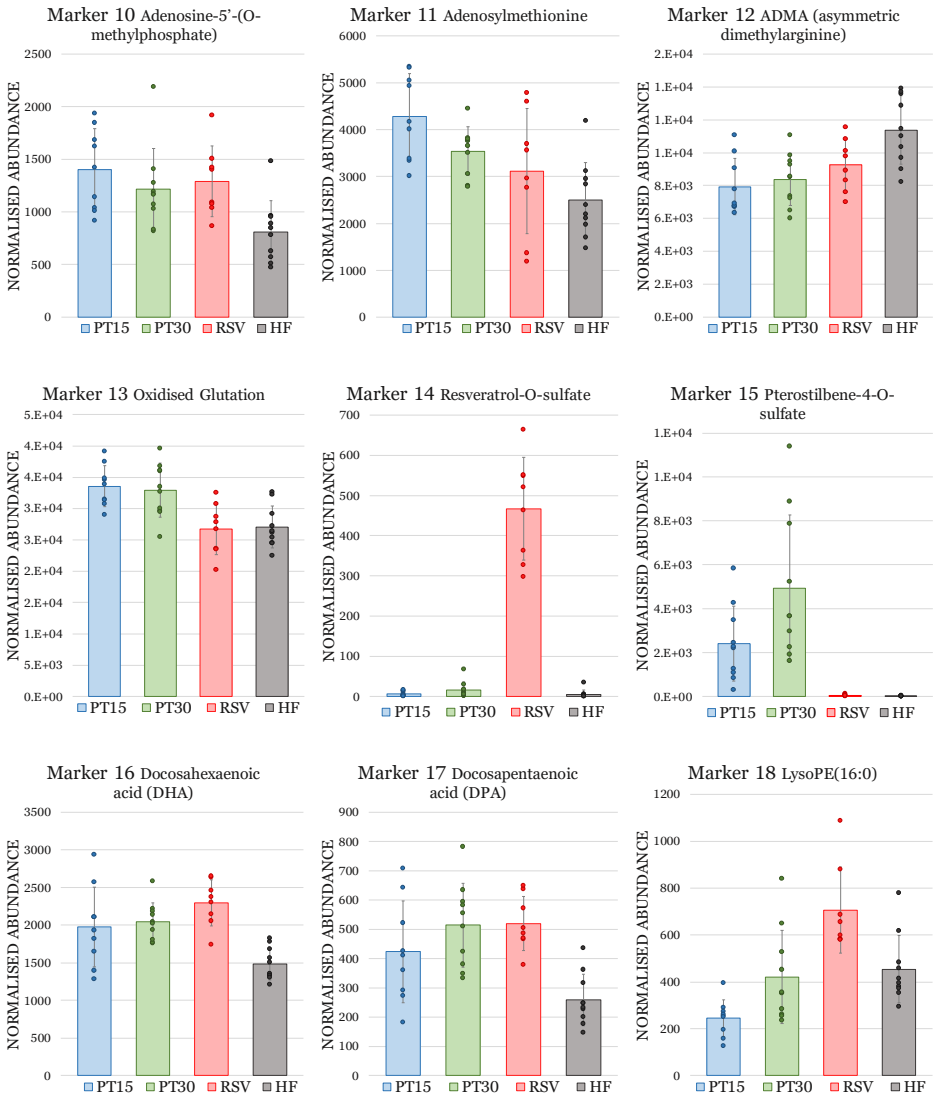


Figure S6. *Comparison of mean of normalised abundance between the different groups samples of markers from 10 to 18: high-fat and high-fructose diet (HF ■), and high-fat and high-fructose diet enhanced with 15 mg/kg/day or 30 mg/kg/day of pterostilbene (PT15 ■ and PT30 ■) or 30 mg/kg/day of resveratrol (RSV30 ■). If the marker is present in more than one analysis, the normalised abundance comes from the marker higher statistical relevance.*

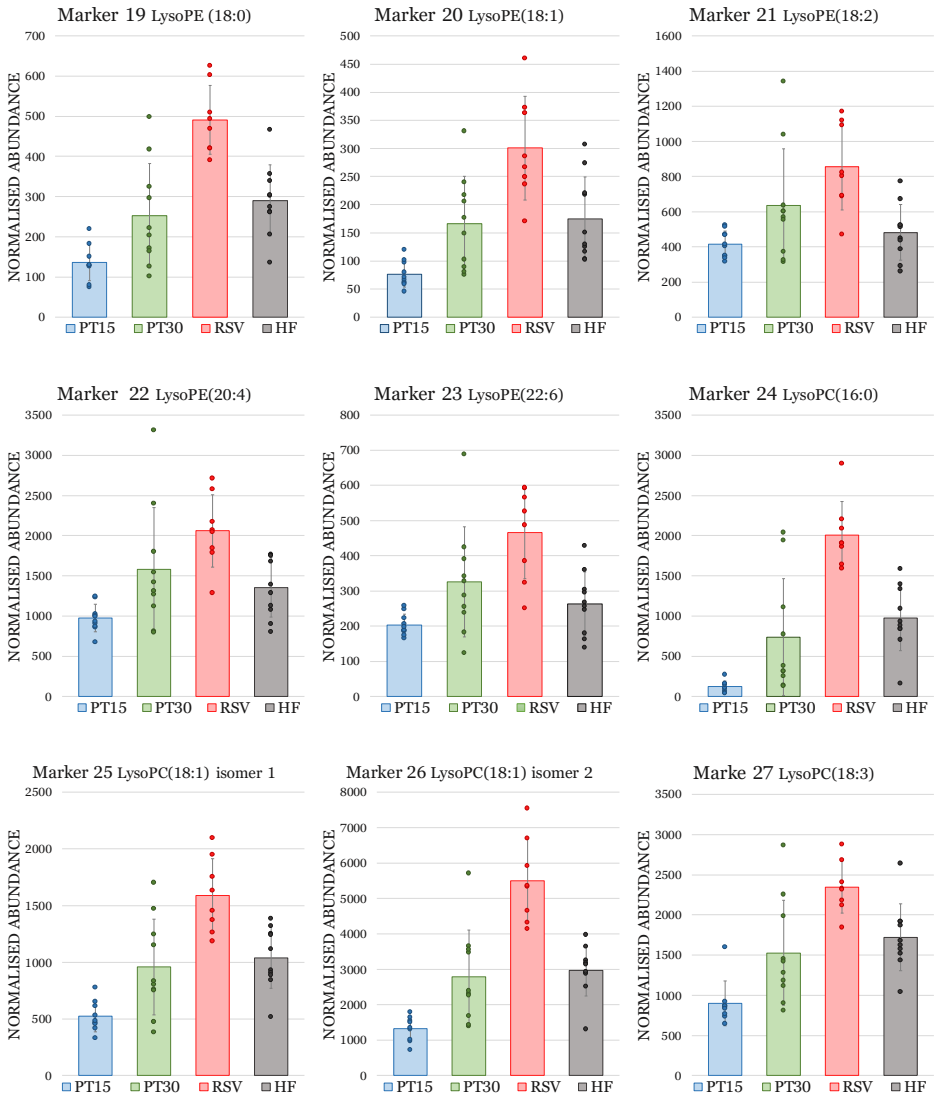


Figure S7. *Comparison of mean of normalised abundance between the different groups samples of markers from 19 to 27: high-fat and high-fructose diet (HF ■), and high-fat and high-fructose diet enhanced with 15 mg/kg/day or 30 mg/kg/day of pterostilbene (PT15 ■ and PT30 ■) or 30 mg/kg/day of resveratrol (RSV30 ■). If the marker is present in more than one analysis, the normalised abundance comes from the marker higher statistical relevance.*

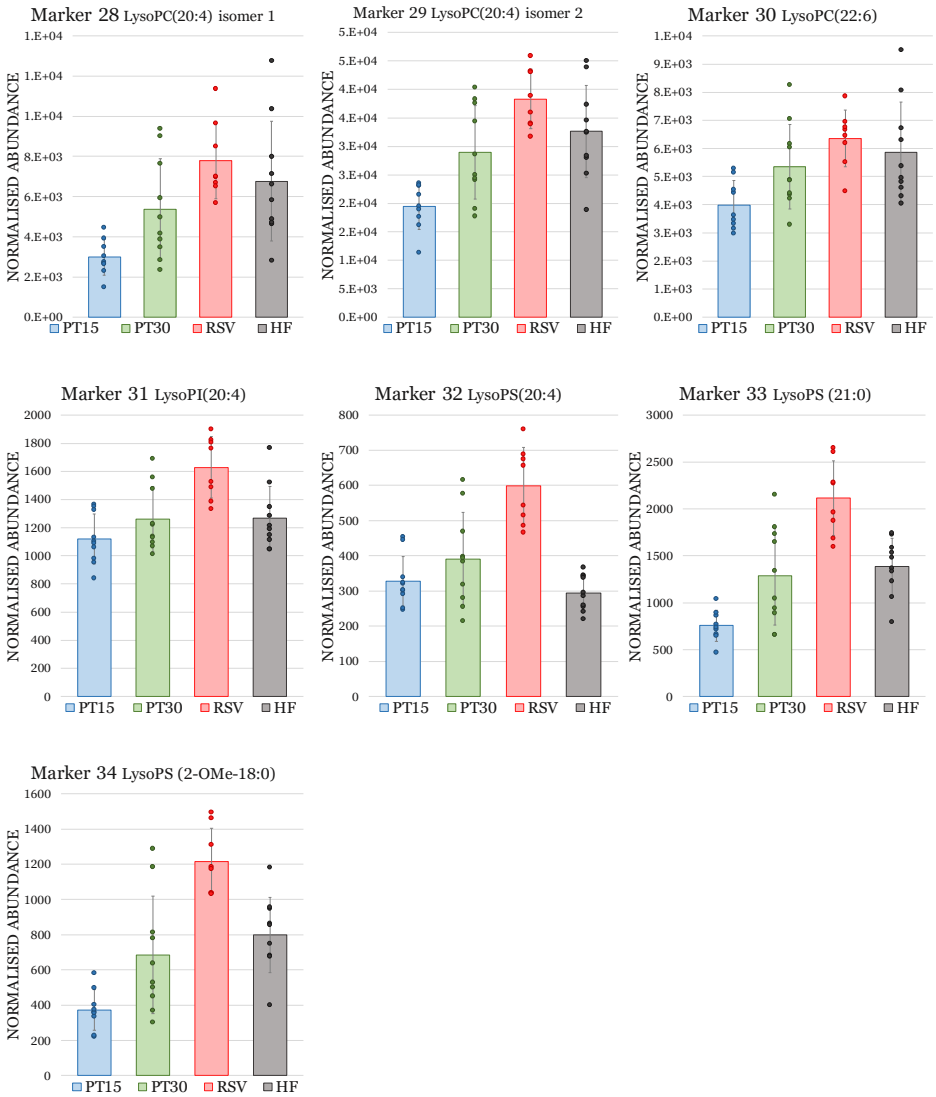


Figure S8. *Comparison of mean of normalised abundance between the different groups samples of markers from 28 to 34: high-fat and high-fructose diet (HF ■), and high-fat and high-fructose diet enhanced with 15 mg/kg/day or 30 mg/kg/day of pterostilbene (PT15 ■ and PT30 ■) or 30 mg/kg/day of resveratrol (RSV30 ■). If the marker is present in more than one analysis, the normalised abundance comes from the marker higher statistical relevance.*

4.3 Discussion of the results

Scientific article IV reflects the great interest of applying untargeted metabolomics approach for the determination of biomarkers in the research area of nutrition and its effects on health. Thirty-four markers related to the supplementation of two known antioxidants, resveratrol and pterostilbene, were tentatively identified in rats with induced liver steatosis by a high-fructose, high-fat diet. The putative relevance of these tentatively elucidated markers as indicators of disease improvement are encouraging. Further studies would investigate whether the supplementation of these compounds would obtain the same results in humans.

In this study, an animal model with rats was used to reduce the bias associated with human trials, since it would allow control of environmental and ingestion conditions, in addition to the fact that the animals are genetically homogeneous. Liver samples were selected to be analysed for being the organ directly affected by the disease under study, liver steatosis. With the naked eye, it was possible to clearly identify the liver samples coming from rat with a standard healthy diet (CC), since the liver presented a clearly more pink and firm appearance compared to the more discoloured and with obvious fat deposits for the other four groups of samples. Therefore, it is not surprising that the main inherent differentiation between the samples highlighted in the first PCA was the differentiation between this group and the liver samples with steatosis. Therefore, and given that the focus of the study was on the effect of resveratrol and pterostilbene in liver steatosis, these samples were discarded. However, this differentiation can be addressed later in further studies.

In this chapter, it is shown the great help for the identification of unknown metabolites that the inclusion of IMS can offer to LC-HRMS coupling, commonly employed in untargeted metabolomics as exposed in the **Scientific article I**. IMS provides additional structure descriptor based on the shape of the molecules. Along with retention time, accurate mass, and fragmentation patterns, CCS values can enhance confidence in identification and reduce false positive hits. Moreover, CCS measurements are more

reproducible values than retention times, more affected by the instrument-to-instrument variation and the characteristics of the sample.

In addition, the development of databases of experimental CCS and the building of CCS prediction tools are being implemented, especially useful when multiple candidates are suggested based on exact mass. In relation to the latter, in this study two CCS prediction tools have been employed, since the effectiveness of these tools depends largely on the set of compounds with which the model has been trained.

On the one hand, the in-house tool of Celma et al. (2021, submitted to Analytical Chemistry) was used for the prediction of the first 17 markers. This tool has been trained with experimental CCS data from about 500 compounds obtained with the same instrument employed in this study (TWIMS based instrument). This set is made up of small molecules (<1000 Da) as organic pollutants, drugs and metabolites, among others. The delta errors reported for this tool is ± 4.05 % for protonated ion and ± 5.86 % for deprotonated ion with 95% confidence interval. All predicted CCS for the markers were below these errors, increasing the confidence in their tentative identification, except for sulphated phase II metabolites with higher error values.

On the other hand, LipidCCS (Zhou et al., 2017) was employed to predict the CCS for the remaining group of markers formed by lysophospholipids. To develop this tool, the CCS data obtained with a DTIMS instrument from about 500 analytical lipid standards were used. They reported deviations between the predicted and empirical data of ~1% in terms of median relative errors (MRE), which would be equivalent to a 50 % confidence interval. Based on that described by Zhou et al (2017), it would be around ±3 % for both protonated and deprotonated molecules in terms of 95 % confidence interval. However, they report that these precisions are only achieved with TWIMS instruments when a lipid-based CCS calibrant are used. When the commonly employed polyAla calibrants are used, the error would probably of above ~5 % with 95 % confidence interval. Even so, it was considered that this model was more suitable for the prediction of the last 17 markers as it has been trained with more similar compounds, obtaining delta errors below 5.4% for protonated molecules and below 4.8% for deprotonated molecules, except for the LysoPS which got higher delta error.

Therefore, with this tool, it was possible to have more confidence about the tentative identity of these markers, without being able to distinguish between the possible isomers for each identification. For each marker, the CCS of all the isomers reported in the LipidMaps compound database were predicted. However, this tool did not provide predicted CCS significantly different from an isomer to another to decide on a single candidate. Zhou et al. (2017) justifies this by the difficulty to find pure standards for isomeric lipid that differ in position geometry (cis/trans) to train the model, and therefore distinguish between isomers.

Several studies related to this work are currently being carried out or scheduled for the near future. On the one hand, the research group with which this collaboration arose is investigating metabolic pathways related to the 34 compounds identified and determining the effects of pterostilbene and resveratrol supplementation on liver steatosis and health based on these results.

Since liver steatosis is due to the accumulation of fats, it was thought that the lipid fingerprint could have significant differences. From our part, the results obtained from an additional untargeted metabolomic focusing on lipidic metabolites (untargeted lipidomics) are also being analysed. In this sense, a second extraction of the same liver samples and LC separation more optimised on lipid analysis were carried out.

In addition, it is planned the application of untargeted metabolomics analysis to plasma samples collected from the same rats, in order to have a global overview of the circulating metabolites throughout the body.

4.4 References

Braconi, D., Bernardini, G., Millucci, L., & Santucci, A. (2018). Foodomics for human health: current status and perspectives. Expert Review of Proteomics, 15(2), 153–164. https://doi.org/10.1080/14789450.2018.1421072

Cifuentes, A. (2013). Foodomics: principles and applications. In A. Cifuentes (Ed.), Foodomics: advanced mass spectrometry in modern food sciences and nutrition (p. 580). John Wiley & Sons, Inc. https://www.wiley.com/en-us/Foodomics%3A+Advanced+Mass+Spectrometry+in+Modern+Food+Science+and+Nutrition-p-9781118169452

Claus, S. P. (2014). Development of personalized functional foods needs metabolic profiling. Current Opinion in Clinical Nutrition and Metabolic Care, 17(6), 567–573. https://doi.org/10.1097/MCO.0000000000000107

Fernández-Ochoa, Á., Leyva-Jiménez, F. J., De la Luz Cádiz-Gurrea, M., Pimentel-Moral, S., & Segura-Carretero, A. (2021). The role of high-resolution analytical techniques in the development of functional foods. International Journal of Molecular Sciences, 22(6), 3220. https://doi.org/10.3390/ijms22063220

García-Cañas, V., Simó, C., León, C., & Cifuentes, A. (2010). Advances in nutrigenomics research: novel and future analytical approaches to investigate the biological activity of natural compounds and food functions. Journal of Pharmaceutical and Biomedical Analysis, 51(2), 290–304. https://doi.org/10.1016/j.jpba.2009.04.019

Gómez-Zorita, S., González-Arceo, M., Trepiana, J., Aguirre, L., Crujeiras, A. B., Irles, E., Segues, N., Bujanda, L., & Portillo, M. P. (2020). Comparative effects of pterostilbene and its parent compound resveratrol on oxidative stress and inflammation in steatohepatitis induced by high-fat high-fructose feeding. Antioxidants, 9(11), 1042. https://doi.org/10.3390/antiox9111042

Mancano, G., Mora-Ortiz, M., & Claus, S. P. (2018). Recent developments in nutrimetabolomics: from food characterisation to disease prevention. Current Opinion in Food Science, 22, 145–152. https://doi.org/10.1016/j.cofs.2018.03.012

Masarone, M., Troisi, J., Aglitti, A., Torre, P., Colucci, A., Dallio, M., Federico, A., Balsano, C., & Persico, M. (2021). Untargeted metabolomics as a diagnostic tool in NAFLD: discrimination of steatosis, steatohepatitis and cirrhosis. Metabolomics, 17(2), 12. https://doi.org/10.1007/s11306-020-01756-1

Picó, C., Serra, F., Rodríguez, A. M., Keijer, J., & Palou, A. (2019). Biomarkers of Nutrition and Health: New Tools for New Approaches. Nutrients, 11(5), 1092. https://doi.org/10.3390/nu11051092

Puiggròs, F., Solà, R., Bladé, C., Salvadó, M. J., & Arola, L. (2011). Nutritional biomarkers and foodomic methodologies for qualitative and quantitative analysis of bioactive ingredients in dietary intervention studies. In Journal of Chromatography A (Vol. 1218, Issue 42, pp. 7399–7414). https://doi.org/10.1016/j.chroma.2011.08.051

Zhou, Z., Tu, J., Xiong, X., Shen, X., & Zhu, Z.-J. (2017). LipidCCS: prediction of collision cross-section values for lipids with high precision to support ion mobility–mass spectrometry-based lipidomics. Analytical Chemistry, 89(17), 9559–9566. https://doi.org/10.1021/acs.analchem.7b02625

Chapter 5.
Intra-specie chemical communication

Chapter 5. Intra specie chemical communication

5.1 Introduction

5.2 Scientific Article V

"Novel sampling strategy for alive animal volatolome extraction combined with GC-MS based untargeted metabolomics: identifying mouse pup pheromones"

Talanta 235 (2021)122786

5.3 Scientific Article VI

"Untargeted metabolomics approach to putative pheromones in mice. Part 2: using UHPLC-IMS-QTOF MS for surface body samples to identify low-volatility chemosignals"

Talanta (2021) Submitted

5.4 Discussion of the results

5.5 References

5.1 Introduction

The interaction is one of the three so-called “vital functions” that any unicellular or multicellular organisms have to perform to be considered a “living being”. From the most solitary to the most gregarious organism, from prokaryotes to animals such as humans; communication intra- and inter-specie are essential for their survival and reproduction, where chemical signalling has a key role in their communication system. The term *semiosignal* refers to all chemical substance involved in the interaction between organisms intra- or inter-specie (Nordlund & Lewis, 1976). Among them, *pheromones* are the compounds, single or in precise ratios of multiple component pheromones, excreted or secreted for intra-specie interaction to trigger a specific reaction, as a behaviour or developmental process on the receptor (Karlson & Lüscher, 1959; Wyatt, 2010).

There are a wide variety of compounds that act as pheromones, including small molecules, peptides or even proteins, and their physicochemical nature is linked to their function. Important chemical properties are their size and polarity, which will determine their volatility in the air and their solubility in water (Brennan, 2009). A pheromone having a rather small size and high volatility allows a rapid dispersion. This also makes these kinds of compounds transitory and they serve to indicate short-term situations as, for example, an indicator of danger or alarm. On the other hand, a pheromone used as territory marker needs a high stability, little volatility and affinity with the surrounding environment.

Pheromone physico-chemical characteristics are also highly related with the releasing in the media and their detection mechanism. Focusing on animals, pheromones release might take advantage of the excretion routes (e.g. urine, faeces and sweat) or by specialised organs or glands for the secretion (Brennan, 2009). In mammals, all chemical signals are detected by the main olfactory system (olfactory epithelium, OE), by the vomeronasal system (VNO) or both. In the case of rodents, they display both chemosensory systems (Fortes-Marco et al., 2013). While the OE detects the mixture of volatiles present in the air, the VNO, present in the roof of the mouth, detects high- and low-volatility compounds. The last ones reach the organ sometimes

mixed in water droplets/spray and thanks to an active mechanism (vomeronasal pumping) (Meredith and O'Connell, 1979).

A comprehensive search in Scopus database has been carried out to investigate the use of untargeted metabolomics together with chromatographic techniques coupled to MS in the field of pheromones study and intra-specie chemical communication. To do so, the search was focused in the last five years and the following keywords were employed: "pheromone", "chemosignal" and "semiosignal"; along with the keywords related to the methodology "untargeted" (and synonyms "fingerprinting", "untargeted" and "non-target"), "metabolomics", "mass spectrometry", "MS" and "HRMS".

The number of research articles obtained have not been very high, although it is a rising approach in this research area, and the determination of new pheromones may have progressed more slowly than other fields. Many of the early strides in pheromone understanding came mainly from biology alone, making predictions of pheromone identity based upon their physiological function and by the screening of commercially available compounds suspected to act as pheromones (Li et al., 2018). Therefore, targeted analytical methods have been the selected approach for most of the studies as terpenoids for in *Camellia sinensis* with GC-EI-Q (Le et al., 2020).

Taking into account that pheromones are compounds for intra-specie communication, they use to be quite specific molecules for each organism and therefore, there is no assurance that it could be present even in other organism of the same taxonomic family. Nevertheless, it is known that there are families of compounds that tend to act as pheromones (Park et al., 2019). Moreover, its presence is not the only factor to provoke a reaction in the receptor, it could be also necessary a certain concentration and/or combination with other molecules in a specific ratio.

The collaboration between chemists and biologists, and the introduction of newer analytical techniques, enabled the development of additional approximations to pheromone research or bioactive compounds characterization by untargeted metabolomics (Butcher, 2017; Hefetz, 2019; Li et al., 2018). Most of the studies have focused on the volatile pheromones

of different organisms and the employment of GC-MS or GC-HRMS instruments to this aim: in yeast (Ljunggren et al., 2019), insects (Esteves et al., 2017) and plants (Hosseini et al., 2021; Zheng et al., 2021). For more polar pheromones, LC-HRMS was used in marine planktonic diatoms (Fiorini et al., 2020) and nematodes (Falcke et al., 2018; Le et al., 2020), where the pheromones are distributed through aqueous media.

Although the introductory chapter of this thesis places special emphasis on the application of the untargeted metabolomic approach along with chromatographic techniques coupled to MS in the discovery of (bio)marker related to food, the workflow explained can be applied to many other fields. The most significant changes in the pheromone study from what was explained in **Scientific article I** is focused in the sampling and sample treatment stage. The crucial issue when trying to analyse a series of compounds with a specific purpose in chemical communication, and which are excreted/secreted under specific biological situations, is to keep or set a determinate biological state in the organism under study or the sampling in the proper time, without introducing a significant change in the chemicals released that could alter the results. These and the fact of having to deal with living organisms can makee difficult to obtain samples, especially for animals. For this reason and taking into account that the type of sample depends on the research aim, the study of pheromones in animals has been mainly performed with easily accessible samples, for example the characterization of aging-related volatiles in mice urine (Fujita et al., 2020); reproductive and pairing status in the urine volatolome of the maned wolf (Jones et al., 2021) or odourants in human sweet related to emotional changes (Smeets et al., 2020). New sample collection techniques are also being developed, as 3D-printed attachment for apocrine sweat collection (Kvasnička et al., 2021).

In this chapter, untargeted metabolomics approach has been applied in the characterization of mice pup pheromones involved in the maternal behaviour. This could be carried out thanks to the close collaboration with the Laboratory of Functional Neuroanatomy from *Universitat Jaume I*, led by the Prof. Dr. Fernando Martínez García, which works on several research lines devoted to the analysis of the neural basis of social and emotional behaviours in rodents mediated by pheromones.

The role of olfaction in maternal care has been widely proven in mammals, and surprisingly, there have not been attempts to identify molecules involved in materno-filial interactions (Bridges, 2008). Some studies suggested the key role of chemical stimuli for maternal care in rodents (Navarro-Moreno et al., 2020), others even proved that induced anosmia, altered chemosensing or olfactory lesions resulted in dramatic changes in the response of dams to pups, from deficient nest maintenance, maternal neglect, abandon and even systematic pup-killing (Hasen & Gammie, 2011; Kimchi et al., 2007; Vandenbergh, 1973). Moreover, mouse dams take care of young pups (even the alien ones) and when they reach the 15-26 days of life, they are progressively rejected until weaning (Numan & Insel, 2003).

Based on last assumption, the working hypothesis was that the pheromones release might vary during the early days of the mice pup and leading to their forced emancipation. Untargeted metabolomic approach was applied for the comparison of the putative pheromones changes between one-week old pups (which elicit an intense postpartum care by females), second week of life and fourth week of life (proximate age of weaning).

As we were interested in the obtention of pup pheromones by repeated sampling in different stages of mice life and do not interfere in their normal development, our priority during the process was keeping the mice pups healthy and comfortable throughout all the studied period. Two complementary studies were developed and carried out simultaneously in the same period of time and with the same animals. In **Scientific article V** an adaptation of the traditional dynamic head space-purge and trap (DHS-P&T) was applied in order to obtain the whole body volatolome and analysed by GC-EI-MS. **Scientific article VI** is focused on lower- volatile pheromones and the sample extracts were analysed by UHPLC-IMS-HRMS.

5.2 Scientific article V

Talanta 235 (2021) 122786

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Novel sampling strategy for alive animal volatolome extraction combined with GC-MS based untargeted metabolomics: Identifying mouse pup pheromones

Leticia Lacalle-Bergeron [a], Rafael Goterris-Cerisuelo [b], Tania Portolés [a,*], Joaquin Beltran [a], Juan Vicente Sancho [a], Cinta Navarro-Moreno [b], Fernando Martinez-Garcia [b]

[a] Environmental and Public Health Analytical Chemistry, Research Institute for Pesticides and Water (IUPA), Universitat Jaume I, Av. Sos Baynat S/N, 12071, Castellón de la Plana, Spain
[b] Laboratory of Functional Neuroanatomy (Unitat Mixta NeuroFun-UV-UJI), Predepartamental Unit of Medicine, Universitat Jaume I, Av. Sos Baynat S/N, 12071, Castellón de la Plana, Spain

ARTICLE INFO

Keywords:
Dynamic headspace
Volatolomics
Untargeted metabolomics
Mouse pheromones
Maternal care
GC-MS

ABSTRACT

In this study, we identify 11 mouse pup volatiles putatively involved in maternal care induction in adult females. For this purpose, we have adapted the dynamic headspace methodology to extract the volatolome of whole alive animals. Untargeted metabolomic methodology was used to compare the volatolome of neonatal (4–6 days) with elder pups until the age of weaning (21–23 days old). Pup volatolome was analyzed by gas chromatography (GC) coupled to single quadrupole mass spectrometry (MS) using automated thermal desorption for sample introduction. After data processing and multivariate statistical analysis, comparison with NIST spectral library allowed identifying compounds secreted preferentially by neonatal pups: di(propylen glycol) methyl ether, 4-nonenal, di(ethylene glycol) monobutyl ether, 2-phenoxyethanol, isomethyl ionone, tridecanal, 1,3-diethylbenzene, 1,2,4,5-tetramethylbenzene, 2-ethyl-p-xylene and tri(propylene glycol) methyl ether. Palmitic acid was enriched in the volatolome of fourth week youngsters compared to neonatal pups. The results demonstrated the great potential of the new sampling procedure combined with GC-MS based untargeted volatolomics to identify volatile pheromones in mammals.

1. Introduction

Animals use chemical senses for inter-individual communication. In many species this includes the excretion or secretion of pheromones, chemicals delivered by an individual that elicit a stereotyped response (either behavioral or neuroendocrine-developmental) in conspecifics [1]. Rodents, which are commonly used in experimental biomedicine, are macrosmatic mammals, e.g. they display two highly developed nasal chemosensory systems [2] for monitoring the presence of relevant chemicals in their environment, the vomeronasal organ (VNO) and the olfactory epithelium (OE).

In the last thirty years, an intense investigation has partially clarified the chemical nature of mouse pheromones mediating a wide variety of social interactions [3]: intersexual attraction, inter-male or maternal aggression; and other behavioural interactions between adult individuals, male avoidance of female youngsters as sexual partners, male effects on female sexual maturation, avoidance of ill conspecifics [4] and pup killing by sexually-naïve males [5]. This has uncovered many molecules putatively involved in social communication in mice, which include volatiles such as farnesenes, thiazolines, heptanones, steroids or small formyl peptides, as well as molecules with lower volatility. Surprisingly, there have not been attempts to identify molecules involved in maternal care, a crucial social behavior that ensures offspring survival and promotes proper neurodevelopment of pups thus facilitating mental and bodily health [6,7]. There is solid evidence indicating a major role of olfaction in pup care: anosmic mothers either eat or abandon their pups [8,9]. Together, these findings suggest a key role of chemical stimuli (mainly volatile odorants) in maternal care, although there is a need of identifying the molecules involved. This requires analyzing the volatiles emitted by the whole body of pups.

The analysis of the emitted volatiles can be done by means of different analytical approaches. Among them, metabolomics, the

* Corresponding author.
E-mail address: tportole@uji.es (T. Portolés).

https://doi.org/10.1016/j.talanta.2021.122786
Received 24 May 2021; Received in revised form 2 August 2021; Accepted 3 August 2021
Available online 4 August 2021

Talanta 235 (2021) 122786

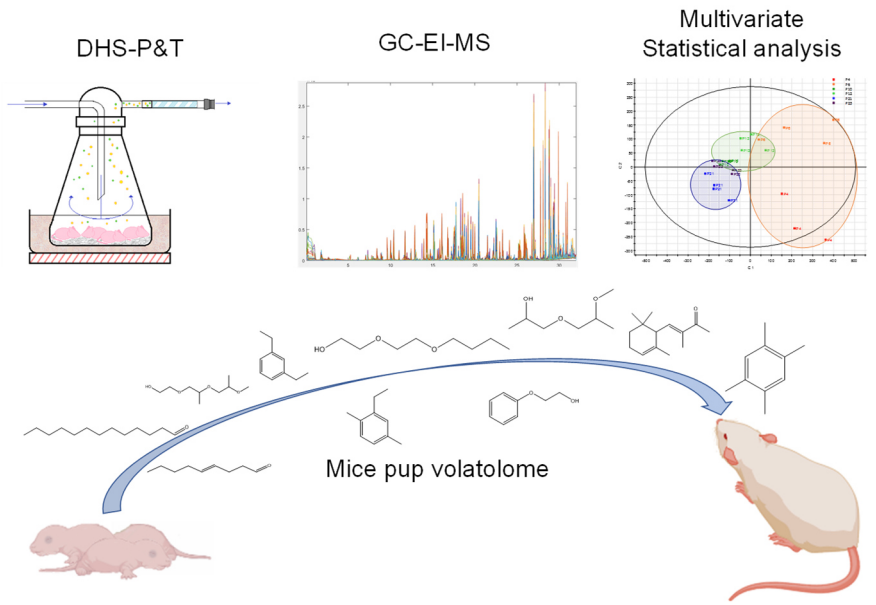
Novel sampling strategy for alive animal volatolome extraction combined with GC-MS based untargeted metabolomics: Identifying mouse pup pheromones

Leticia Lacalle-Bergeron[a], Rafael Goterris-Cerisuelo[b], Tania Portolés[a]*, Joaquin Beltran[a], Juan V. Sancho[a], Cinta Navarro-Moreno[b], Fernando Martinez-Garcia[b]

[a] Enviromental and Public Health Analytical Chemistry, Research Institute for Pesticides and Water (IUPA), University Jaume I, Av. Sos Baynat S/N, 12071 Castellón de la Plana, Spain

[b] Laboratory of Functional Neuroanatomy (Unitat Mixta NeuroFun-UV-UJI), Predepartamental Unit of Medicine, Universitat Jaume I, Av. Sos Baynat S/N, 12071, Castellón de la Plana, Spain

Graphical abstract



Highlights

- Novel dynamic headspace sampling allows whole animal volatolome analysis.
- Untargeted volatolomics is a powerful approach for pheromones determination.
- 11 mouse pups-derived volatiles likely involved in maternal care induction were identified.
- One of them, 1,2,4,5-tetramethylbenzene, is known to activate vomeronasal neurons.

Abstract

In this study, we identify 11 mouse pup volatiles putatively involved in maternal care induction in adult females. For this purpose, we have adapted the dynamic headspace methodology to extract the volatolome of whole alive animals. Untargeted metabolomic methodology was used to compare the volatolome of neonatal (4–6 days) with elder pups until the age of weaning (21–23 days old). Pup volatolome was analysed by gas chromatography (GC) coupled to single quadrupole mass spectrometry (MS) using automated thermal desorption for sample introduction. After data processing and multivariate statistical analysis, comparison with NIST spectral library allowed identifying compounds secreted preferentially by neonatal pups: di(propylen glycol) methyl ether, 4-nonenal, di(ethylene glycol) monobutyl ether, 2-phenoxyethanol, isomethyl ionone, tridecanal, 1,3-diethylbenzene, 1,2,4,5-tetramethylbenzene, 2-ethyl-p-xylene and tri(propylene glycol) methyl ether. Palmitic acid was enriched in the volatolome of fourth week youngsters compared to neonatal pups. The results demonstrated the great potential of the new sampling procedure combined with GC-MS based untargeted volatolomics to identify volatile pheromones in mammals.

Keywords

Dynamic headspace; Volatolomics; Untargeted metabolomics; Mouse pheromones; Maternal care; GC-MS

1. Introduction

Animals use chemical senses for inter-individual communication. In many species this includes the excretion or secretion of pheromones, chemicals delivered by an individual that elicit a stereotyped response (either behavioural or neuroendocrine-developmental) in conspecifics [1]. Rodents, which are commonly used in experimental biomedicine, are macrosmatic mammals, e.g. they display two highly developed nasal chemosensory systems [2] for monitoring the presence of relevant chemicals in their environment, the vomeronasal organ (VNO) and the olfactory epithelium (OE).

In the last thirty years, an intense investigation has partially clarified the chemical nature of mouse pheromones mediating a wide variety of social interactions [3]: intersexual attraction, inter-male or maternal aggression; and other behavioural interactions between adult individuals, male avoidance of female youngsters as sexual partners, male effects on female sexual maturation, avoidance of ill conspecifics [4] and pup killing by sexually-naïve males [5]. This has uncovered many molecules putatively involved in social communication in mice, which include volatiles such as farnesenes, thiazolines, heptanones, steroids or small formyl peptides, as well as molecules with lower volatility. Surprisingly, there have not been attempts to identify molecules involved in maternal care, a crucial social behaviour that ensures offspring survival and promotes proper neurodevelopment of pups thus facilitating mental and bodily health [6,7]. There is solid evidence indicating a major role of olfaction in pup care: anosmic mothers either eat or abandon their pups [8,9]. Together, these findings suggest a key role of chemical stimuli (mainly volatile odorants) in maternal care, although there is a need of identifying the molecules involved. This requires analysing the volatiles emitted by the whole body of pups.

The analysis of the emitted volatiles can be done by means of different analytical approaches. Among them, metabolomics, the "*-omics*" technique focused on the small molecules or metabolites of a biological sample that changes in response to an internal or external alteration (disease, environmental or simply growth) is a good alternative. Although targeted metabolomics has been more commonly used [10], the potential of facing the

study through an unbiased screening global methodology (untargeted approach) seems a good alternative [11,12]. To unravel the biomarkers involved in such changes, statistical tools such as multivariate statistical analysis are employed [13]. Finally, when the study is focused on the volatile or semi-volatile emitted part of the metabolome, the term volatolomics is generally used [14,15].

Gas chromatography coupled to mass spectrometry (GC-MS) is the gold standard for the analysis of the volatile organic compounds (VOCs) as chemical messengers or volatile pheromones [12,14,16]. Compound separation by GC followed by MS detection, traditionally using electron ionization (EI) source, is a sensitive and reproducible combination with a great capability in tentative identification based on the mass spectra search in commercial libraries [12,14,17].

One of the critical aspects when conducting a study of the VOCs that constitute the volatolome is the sampling step. The VOCs are emitted to/from various body fluids and tissues. Their final excretion through physiological fluids and glandular secretion constitutes a matrix of olfactory signals directly related to communication and interaction between animals [14,15,18]. For their extraction from the different biological material, the use of the proper enrichment extraction techniques such as dynamic headspace with sorbent trapping (DHS-P&T) is required. This allows extracting the volatile components from the matrix by dynamic purge of the headspace with, usually, an inert gas and then trapping them in a sorbent. This technique is characterised by its high volatile recovery and pre-concentration factor entailing an enhancement of the sensitivity, along with a good efficiency, low sample manipulation and solventless approach [16,19,20]. Additionally, thanks to the use of a thermal desorption unit (TDU) mounted on a programmed temperature vaporiser injection system the compounds can be directly and efficiently transferred into the GC, further reducing the sample handling and solvent use [14,21].

In our case, however, we are interested in obtaining samples of pup's whole body volatolome and, therefore, animals must be alive and comfortable throughout extraction and the methodology requires repeating sampling in the same animals at different periods of time. This work explores an

adaptation of the traditional methodology by using air instead of an inert gas to remove and preserve the intact volatolome of none-stressed, healthy animals. To the best of our knowledge, this approach has never been used before. To do so, the volatolome of 4-6-day old pups, which elicit intense, dedicated care in postpartum females [22], and the one corresponding to 21–24 day old pups, the approximate age of weaning, when pups are largely rejected by the dam, are compared to identify volatolome changes during pup development.

Therefore, the aim of the work was to apply this novel sampling technique combined with untargeted volatolomics approach based on TDU-GC-MS, to identify putative pup pheromones inducing maternal care in mice.

2. Materials and methods

2.1. Chemicals and reagents

Internal standards (IS) 4,4′-difluorobiphenyl, 4-methyl-2-pentanol and methyl octanoate were purchased from Sigma Aldrich (Germany). An internal standard working solution was prepared with the previously mentioned compounds at 19 μg mL^{-1}, 100 μg mL^{-1} and 25 μg mL^{-1} respectively in hexane. For identity confirmation, the following compounds were purchased: di(propylene glycol) methyl ether, tri(propylene glycol) methyl ether, di(ethylene glycol) monobutyl ether, 2-phenoxyethanol, 1,2,4,5-tetramethylbenzene, 1,3-diethylbenzene, isomethyl ionone and palmitic acid from Sigma Aldrich (St Quentin Fallavier, France); tridecanol from Alfa Aesar (Karlsruhe, Germany); 2-ethyl-p-xylene from abcr GmbH (Karlsruhe, Germany) and 4-nonenal from Ambinter (Orléans, France). Alkane standard solution C8–C20 (Sigma-Aldrich, Germany) was used for Linear Retention Index determination. The organic solvent hexane (trace Analysis quality (AT) GC) was provided by Scharlau (Barcelona, Spain). Tenax® TA glass TD tube, fritted, O.D. 6 mm × 4 mm (i.d.) x L 60 mm, preconditioned, 60–80 mesh, used as traps were purchased from Gerstel (Mülheim an der Ruhr, Germany).

2.2. Animals

In this experiment, a total of n = 4 female mice of the CD1 strain (Janvier Labs, France) of 10 weeks of age were used. Animals were treated throughout according to the European Union Council Directive of June 3rd, 2010 (6106/1/10 REV1). Accordingly, procedures were approved by the Committee of Ethics on Animal Experimentation of the Jaume I University of Castellón where the experiments were performed and, ultimately, by the Valencian *Conselleria d'Agricultura Medi Ambient, Canvi Climàtic i Desenvolupament Rural* (code 2019/VSC/PEA/0049). After mating with an adult male, females were weighted daily to check that they were pregnant and, by day 14–15 of pregnancy, were provided with enough nest material (shredded filter paper) and housed individually in polypropylene cages (145 mm wide, 465 mm long and 215 mm high; Panlab) under controlled temperature (24 ± 2 ºC) and lighting conditions (12 h:12 h; lights ON at 8 a.m.), with *ad libitum* access to food and water. All the precautions were taken to avoid contamination with volatiles coming from the personnel involved in the handling of animals during the experiment (no use of perfumes and/or cosmetics).

2.3. Experimental design for studying volatolome through pup development

Mating was planned so that two females delivered 17 days after the other two. This allowed processing at the same time new-born pups and pups by the age of weaning. When the younger pups were 4 days old and the elder ones were 21-day old, extraction process was started. Volatiles were extracted from animals of the two ages in parallel (see **Table 1**) in two non-consecutive days, thus obtaining volatiles of pups of 4- 6- (first week pups) and 21- 23- days old (fourth week pups). One week afterwards, volatiles were extracted again from the younger pups for two additional days, thus getting data from 10- and 12-day old pups (second week pups).

2.4. Purge and trap extraction procedure from whole alive animals

Volatiles were extracted from groups of sibling pups of both sexes. Pups of the same group were gently removed from the nest, weighed and introduced together in a conic flask with a cotton litter (**Fig. 1**). The number of pups of each group and the volume of the conic flask depended on the postnatal day evaluated: during the 1st week of life, 8 pups in a 150 mL flask (total mass ~26 g), in the 2nd week 6 pups (total mass ~33 g) and a 250 mL flask and finally for the 4th week 3 pups (total mass ~52 g) in a 250 mL flask. The flask was closed with a glass tap with a connection tube for the air entrance, and another as exit that was connected to the sorbent Tenax® TA TDU trap cartridge tube (**Fig. 1**). The sorbent trap was previously spiked with 10 μL of the internal standards working solution for future extraction deviation correction. For 1st and 2nd week pups, whose thermoregulation capability is much reduced, flasks were put on a heated sand bath at 35 – 40 °C during volatile extraction to ensure pup comfort and avoid stress that could modify the volatolome. The DHS-P&T extraction was carried out for 90 min with an air flow of 500 mL min^{-1} induced by a vacuum pump device. Immediately after extraction, pups were returned to the nest with their mother, and a second extraction of the same flasks with the soiled-cotton litter was performed with the same conditions except that the temperature was set at 60 °C. **Table 1** presents the experimental design used in the different extractions. Each day of analysis, a blank with cotton litter (cotton litter blank) and an empty flask blank (air blank) were extracted with the same procedure for each extraction batch. After the extraction process the sorbent trap tubes were desorbed in a thermal desorption unit and the analysis was carried out in a GC-EI-MS analytical platform.

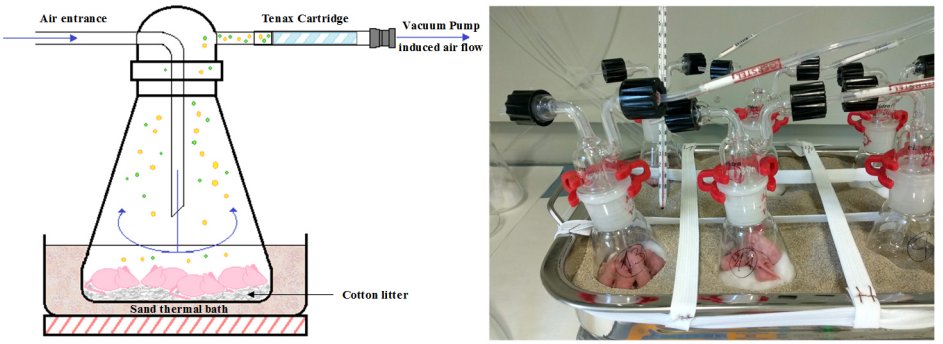


Fig. 1. *Experimental design used for alive mice pup volatolome extraction by purge and trap.*

Table 1. *Experimental conditions of the mice pup volatolome purge and trap extraction.*

Experiment	Nº of replicates	Nº of mice pups	Postnatal week	Postnatal day	Conic flask volume (mL)	Extraction Temperature
1	4	8	1	4	100	35 - 40 ºC
1_nest	4	Pup soiled-cotton litter after experiment 1				60 ºC
2	4	8	1	6	100	35 - 40 ºC
2_nest	4	Pup soiled-cotton litter after experiment 2				60 ºC
3	4	6	2	10	250	35 - 40 ºC
3_nest	4	Pup soiled-cotton litter after experiment 3				60 ºC
4	4	6	2	12	250	35 - 40 ºC
4_nest	4	Pup soiled-cotton litter after experiment 4				60 ºC
5	4	3	4	21	250	Ambient
6	4	3	4	23	250	Ambient

2.5. GC-EI-MS analysis

The gas chromatograph Agilent 6890 Plus Series coupled to a quadrupole mass spectrometer, Agilent 5973 N Mass Selective Detector, with an EI source was used for the mice pup volatolome analysis. The MPS 2 autosampler from Gerstel (Mülheim an der Ruhr, Germany) was employed

and the injection system involved two parts; TDU and CIS 4 PTV. First, the Tenax® TA sorbent tubes were thermally desorbed with the TDU in splitless mode with a desorption program starting at 50 °C with 1 min equilibrium time, and then heated to 260 °C at 12 °C s^{-1} and held for 8 min. The CIS 4 PTV was equipped with Tenax® TA packed liner and the temperature program started at 40 °C during 1 min, followed by a temperature ramp at 12 °C min^{-1} until 260 °C and held for 8 min. The transfer line temperature was set at 260 °C.

The GC separation was carried out on a 30 m × 0.25 mm DB-5MS (0.25 μm film thickness) capillary column (J&W Scientific, Folsom, CA, USA), with helium at a constant flow of 1 mL min^{-1} as carrier gas. The oven temperature program started at 70 °C for 3 min; then increased to 300 °C at 10 °C min^{-1} and held for 9 min (total chromatographic run 35 min).

2.6. Data treatment

The GC-MS data were converted to *netcdf* file format using Chemstation® (Agilent) to perform the data pre-processing with PARADISe (PARAFAC2 based Deconvolution and Identification System) data treatment software. Approximately 200 regions of interest (ROIs) along the chromatogram were manually selected taking into consideration not to leave empty spaces between the intervals and being aware of peak presence when visible in the total ion chromatogram (TIC). The software calculated a model with a maximum of 8 components with 50,000 iterations for each ROI, in order to resolve the underlying and/or overlapping compounds. The models for each ROI were optimised selecting a maximum number of compounds while reaching a background removal, maintaining the model fitting and model consistency over 95 % and avoiding model overfitting. The spectra of each deconvoluted component were automatically compared with the NIST08 (National Institute of Standards and Technology) mass spectral library and each component was tentatively assigned to the best NIST match. Subsequently, a report in *.xls* format was created with the list of compounds and their peak area in each sample. These areas were normalised based on the area of the closest internal standard (IS) and the total mice pup weight of each extraction, to correct the differences due to the instrumental drift and to adjust the emitter volatolome to the body mass of the pups, respectively.

The statistical analysis was performed using EZinfo 3.0 software (Umetrics, Sweden). A pareto-scaling was applied by Ezinfo previous to the evaluation of the samples differences with the multivariate analysis by principal component analysis (PCA), Partial Least Squares – Discriminant Analysis (PLS-DA) and orthogonal PLS-DA (OPLS-DA).

3. Results and discussion

3.1. Mice pup conditions during volatolome sampling

The procedure employed for volatolome extraction apparently did not interfere with pup's growth and comfort. In fact, in many instances, young pups were calmly sleeping during volatolome extraction, thus proving the lack of stress and assuring the intact volatolome. Accordingly, pups were always taken care by the dams when returned to the nest and gained weight according to the standards (**Fig. 2**).

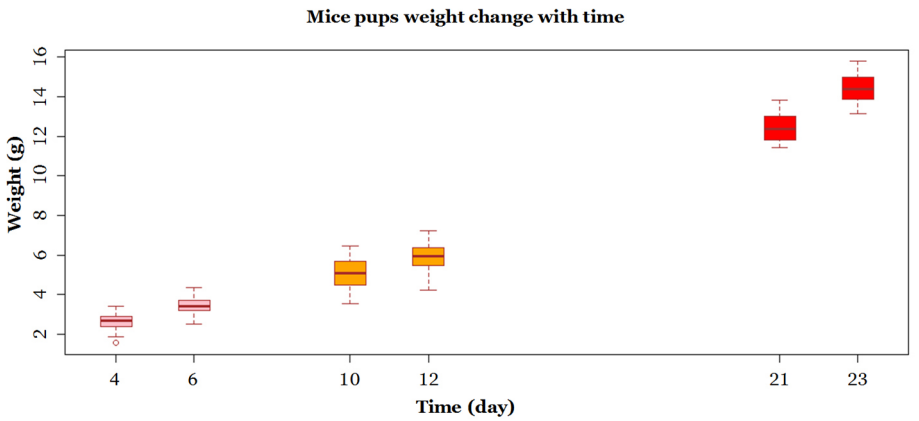


Fig. 2. *Mice pup' weight change with time*

For volatolome extraction, the number of animals was adjusted according to their age (**Table 1**), based on the assumption that the volatolome load is proportional to the total mass of the sampled animals. Despite this, since there is a huge increase in the body weight due to exponential growth of the pups (**Fig. 1**), the total mass sampled was used to normalise the acquired signals.

3.2. Experimental set up for whole volatolome extraction from alive animals

The choice of the volatile extraction technique was based on the previous experience in DHS-P&T in our laboratory [21,23]. The selection of Tenax® TA as the sorbent was based on cited previous works, as good results were obtained for volatiles sampling due to its universality, high retention range and good response to thermal desorption. Nevertheless, the procedure had to be adapted from food matrices to alive mice pups. Firstly, the typical continuous extraction with a flow of an inert gas (as dry N_2) was unfeasible to preserve the mice pups' life. The carrier gas was substituted by ambient air, whose entrance was forced with a vacuum pump device and the flow was set to approximately 500 mL min^{-1} draw-in air. Although the purging gas used was not inert, it was assumed that the possible oxidations occurred in the volatolome with the ambient air were those that naturally occur and, therefore, the markers detected would be those that the mouse dams actually perceived. Moreover, in order to rule out compounds coming from the air and/or the cotton litter, blanks were also extracted (empty conic flask and conic flask with cotton litter, respectively).

Secondly, it is usual to increase temperature to promote the release of the volatile components from the matrix. In the case of the extractions with the mice pup, increasing the temperature too much could stress pups thus possibly causing unwanted changes in the volatolome. Therefore, the sand bath was set at 40 °C in mice pup volatolome extractions of the first and second week, to reach a temperature of 37–39 °C within the flasks, which corresponds to the comfort temperature for the pups. Extractions from fourth week pups, which are able to thermoregulate, were carried out at room temperature, 24 °C during the experiments. When extracting the volatile compounds that remain in the soiled-cotton litter after the 90 min period of the first volatolome extraction, the temperature was raised to 60 °C to further facilitate release of volatile compounds. This second experiment was carried out in order to obtain complementary information about the behaviour of pup pheromones. Specifically, the experiment was aimed to check if a portion of the pheromones emitted by pups is retained in the nest where pups are living. Nevertheless, this was done only in the two first weeks of study because after

the extraction, fourth week mice pup used to defecate abundantly and, therefore, cotton bedding of these pups was discarded because it can be contaminated by bacteria-derived volatiles.

In chromatography-MS-based metabolomics, it is a common procedure to monitor the quality of the results and the analytical process by means of quality controls (QCs, a homogeneous pool of all samples). Moreover, for standardizing the data acquisition process and minimize the bias, it is advisable to randomize as much as possible the samples in the sequence of analysis [12]. Unfortunately, none of these methodologies could be applied as the extractions were carried out as mice pup grew up, therefore there was no way to randomize the samples and, obviously, to make a QC. Instead, each sorbent trap was spiked with a mix of internal standards prior to the extraction process to ensure stability and correct the possible instrumental drift as well as extraction deviation.

3.3. Data processing

The data processing started, as indicated above, with the conversion of GC-MS data to a machine-independent format (*.cdf*) using Chemstation® (Agilent Technologies).

PARADISe [24,25] was the processing software chosen based on the previous satisfactory experience in our laboratory for the peak picking and retention time alignment in VOCs analysis [21,26].

All sample data were processed simultaneously, and using the interactive visualization of the software, the composite of the total ion chromatograms (TIC) were divided into 199 ROIs. Each interval was then individually PARAFAC2 modelled [27,28], which allows the peak deconvolution based on the intensity and the spectra of the signals. Following Khakimov et al. (2012) [29] recommendations, the validation of each model was conducted. The software calculates models for each ROI from one to eight components, and the optimal number was decided attending a good model fitting and core consistency (both over 95 %), low residuals, noise removal and avoiding model overfitting. Only the well resolved peaks with a robust NIST match were selected for statistical processing and those that represents baseline and artefact peak as column bleeding were eliminated.

The optimisation of the models from the 199 intervals of the GC-MS raw data resulted into approximately 173 components tentatively identified and recorded with their peak area in a final report as *.xls* data table.

The capabilities of PARADISe to resolve complex data system with little user interaction, with a data analysis procedure transparent, simple and time effective, allows to obtain consistent data matrix that facilitates the following statistical steps. For further information, this software has been explained in detail in our previous work of Sales et al. (2019) [21].

To correct the possible instrumental deviation, raw data were processed dividing the peak area of each compound by the peak area of the nearest IS in each sample. Then, the data were also corrected by the total weight of the mice pup in each extraction. Prior to multivariate analysis, a pareto-scaling to the obtained peak data was also applied.

3.4. Multivariate statistical analysis

Firstly, the unbiased PCA was performed as exploratory analysis of the data obtained from the volatolome extracted from the pups of different ages. **Fig. 3** shows the evolution of the PCA loadings plot after the data corrections mentioned before. Firstly, the main information that can be extracted from the PCA applied to the non-corrected data (**Fig. 3a**) is that the blanks extraction, both from the empty flask (blank_syst) or the flask containing only clean cotton litter (blank_litter), are grouped and separated from the other groups of samples, evidencing that they are significantly different to the mice pup volatolome extractions. Once this is checked, blank data were removed for subsequent statistical analysis.

Furthermore, it can be observed an intrinsic separation of the samples by post-natal day and also by week, from left to right (COMP.1) from the younger to the oldest mice pup. Since the first principal component of a set of features is the normalised linear combination of the features, this indicates a global, general increase in volatolome size with age. Seven principal components were necessary to explain the 79 % of the explained variance, where first and second component only explain the 29 % and 15 % of the variance, respectively. After the IS correction (**Fig. 3b**) the variance explained was 83 % with five components, of which 42 % corresponds to

Comp.1 and 15 % to Comp.2. The separation of the groups by day or by week is not significantly improved by the correction. This could be due to the fact that the instrumental drift is not as significant as in other longer analysis runs, and therefore this correction has not a strong impact on the data. Nevertheless, the statistical procedure was continued with the corrected data. Finally, in **Fig. 3c** it can be observed the effect of correction by the weight. Five components were enough to explain the 82 % of the variance, 43 % by Comp.1 and 15 % by Comp.2. Comp.1 explains the separation by age, more evident if the samples are grouped by week, from right to left from the earlier week to the last week of life analysed; while Comp.2 apparently finds significant differences between day 4 and day 6 of mice pup life. A possible reason why volatolome could be affected in this way is that during this two-day period the mice pup began to develop fur [30].

The same processing workflow was applied to the data of soiled-cotton litter extraction at 60 °C performed after the mice pup volatolome extraction. The PCAs obtained after correction by internal standard and mice pup weight showed that only three components were necessary to explain the 76 % of the total variance and it can be observed a separation between the samples by days of life and by week (**Fig. S1**).

Subsequently, after the primary examination of the data with PCA, the partial least squares discriminant analysis (PLS-DA) was applied. This supervised multivariate statistical analysis considers additional information about the samples to try to reach a target grouping. Based on the information extracted from the PCA, it was considered that it might be more interesting to attempt the grouping by weeks. The PLS-DA obtained (**Fig. 4** and its corresponding loading plot **Fig. S2**) reach a successful classification of the samples of mice pup extracted volatolome with 97 % of the variance explained and 87 % of the predicted with six latent variables. The groups (weeks) are separated along the first latent variable, which explains the 43 % of the total variance. Moreover, there is also a separation along the second latent variable (11 % of the explained variance) related to the second week of life.

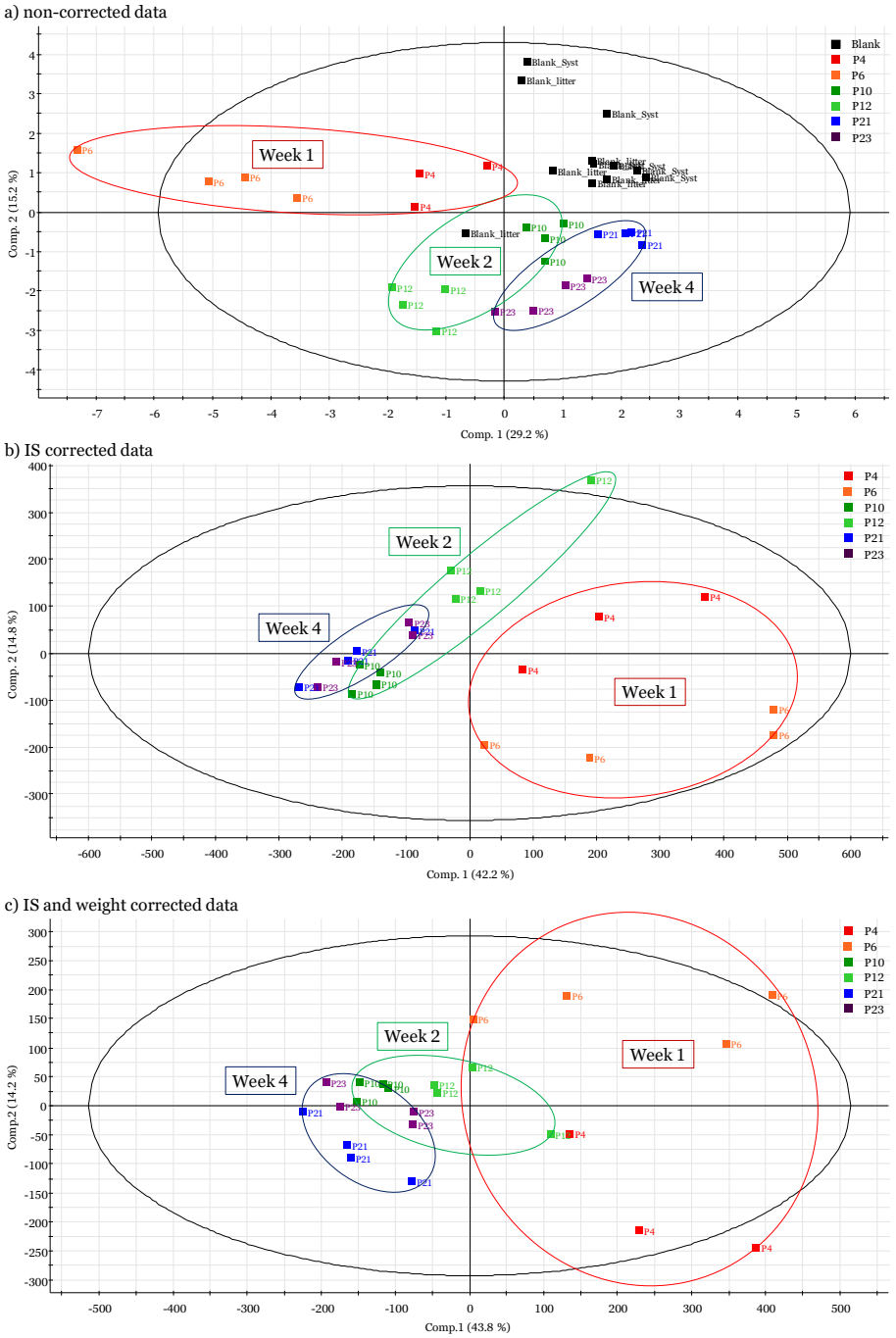


Fig. 3. *PCA score plots in the plane Comp.1 vs Comp.2 of the samples of mice pup volatolome extraction: (**a**) the non-corrected data with blank samples; (**b**) data corrected the nearest internal standard and (**c**) data corrected by the nearest IS and subsequently by the total weight of mice pup in each extraction.*

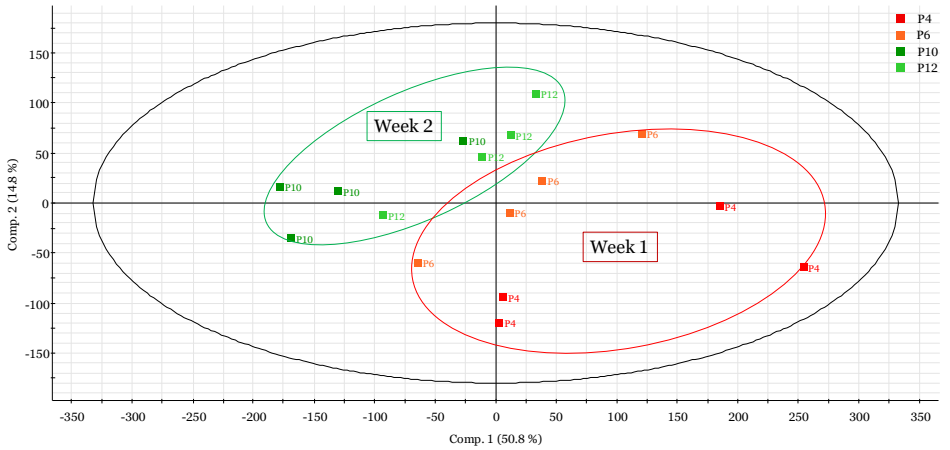


Fig. S1. *PCA score in the plane Comp.1 vs Comp.2 of the remaining volatolome in the soiled-cotton litter extraction at 60 ºC data, corrected by the nearest IS and subsequently by the total weight of mice pup in each extraction.*

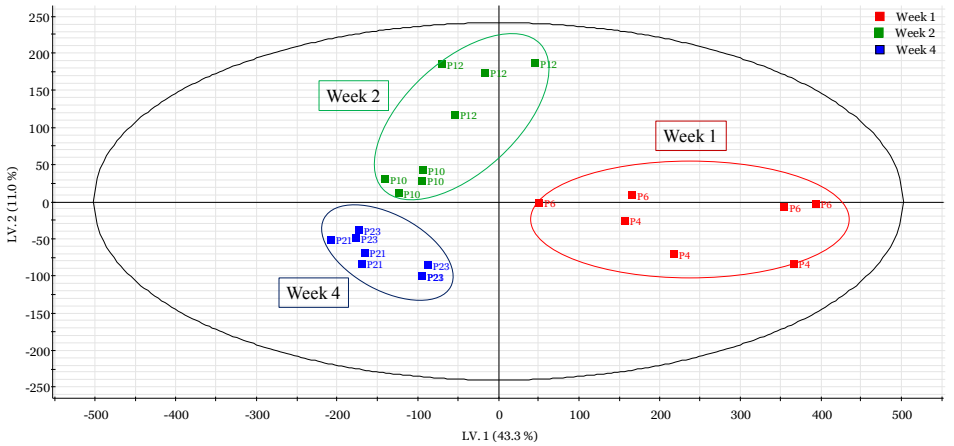


Fig. 4. *PLS-DA in the LV.1 vs LV.2 plane score plot of the samples of mice pup volatolome extraction data (internal standard and mice pup the total weight corrected) grouped by weeks of life. From the 97 % of the variance explained in this model 43 % and 11 % corresponds to LV.1 and LV.2 respectively.*

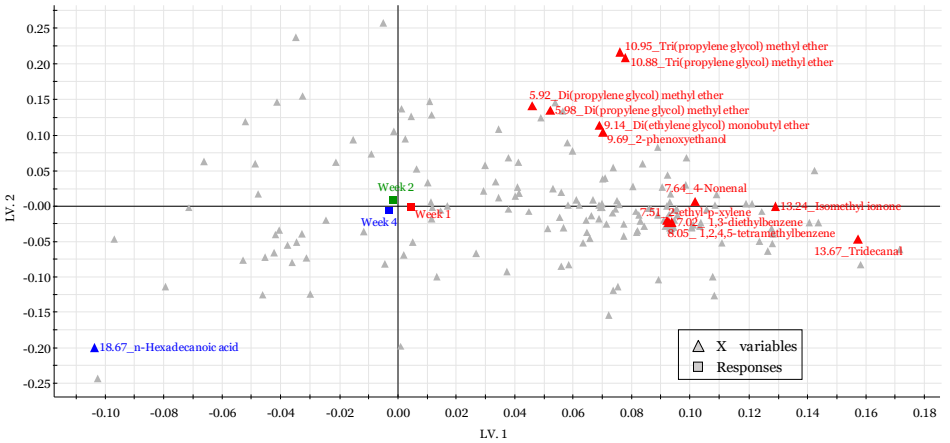


Fig. S2. *Corresponding Loading Plot for the PLS-DA model obtained from the samples of direct mice pup volatolome extraction data (internal standard and mice pup the total weight corrected) grouped by weeks of life.*

Since the objective was to focus on those volatiles that are present at high concentrations in the volatolome of young pups, an OPLS-DA was applied facing the logically more dissimilar of the mice volatolome extraction groups, week 1 vs week 4. The S-plot from the OPLS-DA permitted to highlight the markers more relevant for this differentiation (**Fig. 5**): In this plot the markers are distributed according to their discriminatory power between the two groups selected, being the most discriminant those with a *p(corr)* closer to 1. It was used as a threshold to select the most relevant marker a *p(corr)* > 0.8, obtaining 42 compounds which had a higher abundance in week 1 and 2 than in week 4, and 3 compounds with higher abundance in the week 4. It was studied the presence of these features in the blank samples (blank_syst and blank_litter), and 11 out of 45 were discarded as their signal was relevant enough and cast doubt on their validity as markers.

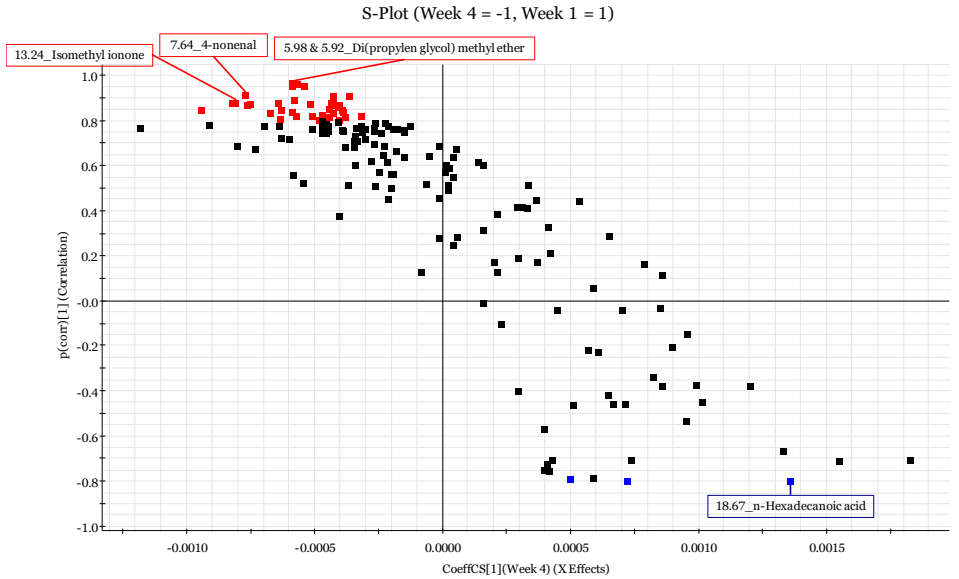


Fig. 5. *OPLS-DA S-Plot for individual biomarker highlighting. Markers with a p(corr) > 0.8 are coloured in red and blue with higher presence in the week 1 and week 4 respectively.*

3.5. Elucidation process

Putative identifications of the relevant compounds selected form the S-plot were reviewed. These tentative identifications were obtained thanks to the automatic comparison between the deconvoluted spectra and the NIST EI spectral library (NIST08 version) performed by PARADISe software. Additionally, the Linear Retention Indices (LRIs) were calculated for each compound using a C7–C20 alkane mixture. Those features with NIST match factor lower than 700 and/or a LRIs match with the NIST library below ±20 were not identified, remaining a total of 24 markers with a reliable tentative identification.

Finally, 15 markers corresponding to 13 different compounds were selected to be purchased according to their discrimination power and their availability. Following the criteria of our laboratory and the *Chemical Analysis Working Group* (CAGW) *Metabolomics Standards Initiative* (MSI) [31], their identities were confirmed by the comparison of spectra and retention time with the reference standard injected under the identical

sample analysis conditions. Two putative markers did not match with the retention time of reference standard, and therefore were discarded. **Table 2** shows the results of the 13 markers finally identified, with their retention time, molecular formula, the molecular ion, the NIST match and the experimental and theoretical LRI. In the case of the two peaks corresponding to Tri(propylene glycol) methyl ether, no theoretical LRIs were found, but their identities were confirmed with the injection of the commercially available mixture of isomers. **Fig. S3** shows intensity variations across the samples for each of these markers using variable trend plots.

3.6. Putative pup volatile pheromones

As shown in **Table 2** and **Fig. S3**, we have identified several volatile compounds that are secreted by pups of the 1st and 2nd weeks of age, when pups elicit intense care in adult females and motivated maternal behaviour in lactating dams [32], but their secretion decreases as the pups get older and are very scarce in the volatolome of 4-week youngsters (the age of weaning). Therefore, these compounds can possibly act as chemosignals in the context of mother-pup interactions, thus maybe qualifying as pup-derived maternal behaviour-inducing pheromones.

Among the compounds identified in the volatolome of 1st and 2nd week old pups, there are three glycol ethers (di(ethylen glycol) methyl ether, di(propylen glycol) methyl ether and tri(propylen glycol) methyl ether). We have not found previous reports of these compounds being direct metabolites in vertebrates or invertebrates, even if they are commonly used in industry (as solvents and hydraulic fluids) and, as pollutants, their derived metabolites have been analysed in rodents [33] in the context of their possible toxicity.

On the contrary, 4-nonenal has been reported as a pheromone secreted by females of *Drosophila* [34] and also of the stink bug *Acrosternum aseadum* [35]. As far as we know, no previous reports indicate its role in chemical communication in vertebrates.

Table 2. *GC/MS measurements for the identified pup-derived volatile markers.*

P[Corr] value	MARKER	t_R (min)	Molecular formula	M+• (*m/z*)	NIST match	LRI Experimental [a]	LRI Reported in NIST library [a]
0.96	Isomer 1 Di(propylen glycol) methyl ether	5.98	$C_7H_{16}O_3$	148 [c]	907	1000	981
0.95	Isomer 2	5.92			894	997	
0.91	4-nonenal	7.64	$C_9H_{16}O$	140	893	1095	1096
0.90	Di(ethylene glycol) monobutyl ether	9.14	$C_8H_{18}O_3$	162 [c]	860	1189	1189
0.88	2-phenoxyethanol	9.69	$C_8H_{10}O_2$	138	869	1222	1221
0.88	Isomethyl ionone	13.24	$C_{14}H_{22}O$	206	702	1477	1473
0.88	Tridecanal	13.67	$C_{13}H_{26}O$	198 [c]	903	1511	1513
0.87	1,3-diethylbenzene	7.02	$C_{10}H_{14}$	134	802	1060	1055
0.87	1,2,4,5-tetramethybezene	8.05	$C_{10}H_{14}$	134	772	1120	1116
0.87	2-ethyl-p-xylene	7.51	$C_{10}H_{14}$	134	880	1088	1093
0.87	Isomer 1 Tri(propylene glycol) methyl ether	10.88	$C_{10}H_{22}O_4$	206 [c]	928	1302	[b]
0.87	Isomer 2	10.95			797	1307	
-0.87	n-Hexadecanoic acid (palmitic acid)	18.67	$C_{16}H_{32}O_2$	256	914	1961	1963

[a] *The Linear Retention Index (LRI) were obtained for each compound from NIST Library (https://webbook.nist.gov/) according to the most similar column and chromatographic conditions.*
[a] *There are not reported LRI in NIST library for this compound.*
[c] *Not present in the EI spectra.*

Another of the 1st and 2nd week pup secreted volatiles, 2-phenoxyethanol, was detected in the secretion of the chin gland of adult male rabbits [36], where its concentration rises when the male becomes dominant. However, since this compound is commonly used in cosmetics to fix odorants, it has been postulated that 2-phenoxyethanol may also subserve a similar function in rabbit chemical communication, so that adding it to chin secretion would facilitate odours of dominant males to persist in the environment and not dissipate. Whether a similar function occurs in mouse pups, requires further investigation.

The volatolome of 1-week mouse pups is also enriched in an interesting compound, 1,2,4,5-tetramethylbenzene or durene, the presence of which decreases already in the second week of age. Two other similar compounds (2-ethyl-p-xylene and 1,3-diethylbenzene) show a similar profile of secretion during pup maturation (**Fig. S3**). In a pioneer study, Sam et al. (2001) [37] used Ca^{++} imaging to identify compounds that activate specifically mouse vomeronasal neurons *in vitro*. They tested several substances previously

suggested to be pheromones, but also mixtures of odorants for which a pheromonal role had not been proposed. Among them, they included camphoric odorants, of which, only durene showed a brisk, specific activation of some vomeronasal neurons. Twenty years afterwards, our work shows that this compound is naturally present in the volatolome of pups, and very enriched in 1-week pups, thus suggesting a possible role for this and related molecules, as pup pheromones eliciting maternal behaviour in dams. This will be tested in the near future.

As compared to mouse youngsters at the time of weaning, neonatal pups also secrete several other compounds for which we have not found previous reference in the literature of acting as chemosignals. This includes 2-ethyl-p-xylene, isomethyl ionone and tridecanal. Some of these compounds are plant-related odorants detected by the olfactory system of insects (1,3-diethylbenzene [38] and tridecanal [39] sometimes used in cosmetics and perfume industry but, to the best of our knowledge, there are no previous reports of they being involved in animal communication.

Finally, there is a single molecule identified as enriched in the volatolome of 4-week youngsters as compared to neonatal pups, hexadecanoic acid (palmitic acid). This fatty acid is very common in many vegetal oils, and it cannot be discarded that its appearance in the volatolome of mouse pups at weaning be related to autonomous feeding of standard food chops, maybe enriched in this nutrient. Nonetheless, its possible role in materno-filial communication should be explored in future investigations.

Fig. S4 shows the variation in the intensity obtained across the weeks and the comparison with the obtained ones in the soiled-cotton litter extraction. It can be observed that only part of the volatolome is retained in the cotton bedding, and even not retained in some case as for isomethyl ionone and 1,3-dimethylbenzene. This suggest whereas the remaining compounds may reach the bedding, as for example if they are excreted in the urine, these two compounds may be released using a different way of secretion/excretion. As hexanoic acid was not found either in the cotton litter in the first two weeks, and the cotton bedding of the fourth week was not analysed.

4. Conclusions

This work has proven the usefulness and validity of a novel sampling procedure for volatolome extraction of whole, alive mouse pups. Its combination with untargeted metabolomics GC-MS approach has allowed identifying putative pheromones involved in maternal behavior. Investigation on the volatolome of mouse pups from first week (neonatal, receiving maternal care) to fourth week of age (time of weaning) has shown a more rich and abundant volatolome in neonatal pups (receiving more maternal care), supporting previous evidence that suggests a role of chemical communication in maternal behavior. Several of the compounds of neonatal pups are virtually absent in 4-week youngsters and are candidates to pup pheromones eliciting maternal care in adult females. The use of the software PARADISe in combination with NIST spectral libraries have allowed a robust peak detection and an effective tentative identification. Eleven of these putative pheromones identities have been confirmed with the commercially available standard, one of which, 1,2,4,5-tetramethylbenzene, has previously been shown to activate specifically sensory cells of the vomeronasal organ of adult male and female mice. On the other hand, we have identified a single compound, palmitic acid, which is enriched in the volatolome of 4-week youngsters as compared to neonatal pups, although it is not clear yet whether it is related with the change in diet associated to weaning. Further research is needed to clarify the putative role of these compounds in materno-filial communication and the regulation of maternal behaviour.

Deontological

These experiments were performed throughout following the European Union Council Directive of June 3rd, 2010 (6106/1/10 REV1). Accordingly, procedures were approved by the Committee of Ethics on Animal Experimentation of the Jaume I University of Castellón where the experiments were performed and, ultimately, by the Valencian *Conselleria d'Agricultura Medi Ambient, Canvi Climàtic i Desenvolupament Rural* (code 2019/VSC/PEA/0049).

Credit author statement

Leticia Lacalle-Bergeron: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Funding acquisition. **Rafael Goterris-Cerisuelo**: Conceptualization, Methodology, Formal analysis, Investigation, Visualization. **Tania Portolés**: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Joaquim Beltrán**: Conceptualization, Methodology, Validation, Investigation, Resources, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Juan Vicente Sancho**: Conceptualization, Methodology, Validation, Investigation, Resources, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Cinta Navarro-Moreno**: Conceptualization, Investigation, Resources, Visualization **Fernando Martínez-García**: Conceptualization, Methodology, Validation, Investigation, Resources, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

L Lacalle-Bergeron acknowledges the financial support of Universitat Jaume I, Spain for his pre-doctoral grant (UJI 19I001/03). T. Portoles acknowledges Ramon y Cajal Program from the Ministry of Economy and Competitiveness, Spain (RYC-2017-22525) for funding her research. The Research Institute for Pesticides and Water (IUPA) authors acknowledge the financial support of Generalitat Valenciana, as research group of excellence PROMETEO/2019/040 and Universitat Jaume I de Castelló (UJI-B2020-25). The Laboratory of Functional Neuroanatomy (Unitat Mixta NeuroFun-UV-UJI) authors acknowledge for financial support of Generalitat

Valenciana, as research group of excellence PROMETEO/2017/078, the Spanish Ministry of Science and Innovation (BFU2016-77691-C2; PID2019-107322 GB-C21) and Universitat Jaume I de Castelló (UJI-B2016-45).

References

[1] P. Karlson, M. Lüscher, 'Pheromones': a new term for a class of biologically active substances, Nature, 183 (1959), pp. 55-56, 10.1038/183055a0

[2] L. Fortes-Marco, E. Lanuza, F. Martinez-Garcia, Of Pheromones, Kairomones, What receptors mediate innate emotional responses?, Anat. Rec., 296 (2013), pp. 1346-1363, 10.1002/ar.22745

[3] L. Stowers, T.-H. Kuo, Mammalian pheromones: emerging properties and mechanisms of detection, Curr. Opin. Neurobiol., 34 (2015), pp. 103-109, 10.1016/j.conb.2015.02.005

[4] B. Bufe, Y. Teuchert, A. Schmid, M. Pyrski, A. Pérez-Gómez, J. Eisenbeis, T. Timm, T. Ishii, G. Lochnit, M. Bischoff, P. Mombaerts, T. Leinders-Zufall, F. Zufall, Bacterial MgrB peptide activates chemoreceptor Fpr3 in mouse accessory olfactory system and drives avoidance behaviour, Nat. Commun., 10 (2019), p. 4889, 10.1038/s41467-019-12842-x

[5] Y. Isogai, Z. Wu, M.I. Love, M.H.-Y. Ahn, D. Bambah-Mukku, V. Hua, K. Farrell, C. Dulac, Multisensory logic of infant-directed aggression by males, Cell, 175 (2018), pp. 1827-1841, 10.1016/j.cell.2018.11.032

[6] M. Numan, T.R. Insel, The Neurobiology of Parental Behavior (first ed.), Springer-Verlag New York, new york (2003), 10.1007/b97533

[7] R.S. Bridges, Neurobiology of the Parental Brain, Elsevier (2008), 10.1016/B978-0-12-374285-8.X0001-7

[8] R. Gandelman, M.X. Zarrow, V.H. Denenberg, M. Myers, Olfactory bulb removal eliminates maternal behavior in the mouse, Science, 171 (1971), pp. 210-211, 10.1126/science.171.3967.210

[9] Z. Wang, D.R. Storm, Maternal behavior is impaired in female mice lacking type 3 adenylyl cyclase, Neuropsychopharmacology, 36 (2011), pp. 772-781, 10.1038/npp.2010.211

[10] S. Ihara, K. Yoshikawa, K. Touhara, Chemosensory signals and their receptors in the olfactory neural system, Neuroscience, 254 (2013), pp. 45-60, 10.1016/j.neuroscience.2013.08.063

[11] S. Moco, J. Vervoort, S. Moco, R.J. Bino, R.C.H. De Vos, R. Bino, Metabolomics technologies and metabolite identification, TrAC Trends Anal. Chem. (Reference Ed.), 26 (2007), pp. 855-866, 10.1016/j.trac.2007.08.003

[12] A. Garcia, C. Barbas, Gas chromatography-mass spectrometry(GC-MS)-based metabolomics, T.O. Metz (Ed.), Metab. Profiling, Methods Protoc. (2011), pp. 191-204, 10.1007/978-1-61737-985-7

[13] B. Worley, R. Powers, Multivariate analysis in metabolomics, Curr. Metabolomics., 1 (2012), pp. 92-107, 10.2174/2213235X130108

[14] S. Giannoukos, A. Agapiou, B. Brkić, S. Taylor, Volatolomics: a broad area of experimentation, J. Chromatogr., B, 1105 (2019), pp. 136-147, 10.1016/j.jchromb.2018.12.015

[15] Y.Y. Broza, P. Mochalski, V. Ruzsanyi, A. Amann, H. Haick, Hybrid volatolomics and disease detection, Angew. Chem. Int. Ed., 54 (2015), pp. 11036-11048, 10.1002/anie.201500153

[16] M.V. Novotny, H.A. Soini, Analysis of Volatile Mouse Pheromones by Gas Chromatography Mass Spectrometry, (2013), pp. 29-45, 10.1007/978-1-62703-619-1_3

[17] W.B. Dunn, D.I. Ellis, Metabolomics: current analytical platforms and methodologies, TrAC Trends Anal. Chem. (Reference Ed.), 24 (2005), pp. 285-294, 10.1016/j.trac.2004.11.021

[18] B. de Lacy Costello, A. Amann, H. Al-Kateb, C. Flynn, W. Filipiak, T. Khalid, D. Osborne, N.M. Ratcliffe, A review of the volatiles from the healthy human body, J. Breath Res., 8 (2014), Article 014001, 10.1088/1752-7155/8/1/014001

[19] S. Kakuta, S. Nishiumi, M. Yoshida, E. Fukusaki, T. Bamba, Profiling of volatile compounds in APCMin/+ mice blood by dynamic headspace extraction and gas chromatography/mass spectrometry, J. Chromatogr., B, 1003 (2015), pp. 35-40, 10.1016/j.jchromb.2015.09.002

[20] A.C. Soria, M.J. García-Sarrió, A.I. Ruiz-Matute, M.L. Sanz, Headspace techniques for volatile sampling, Compr. Anal. Chem., Elsevier B.V. (2017), pp. 255-278, 10.1016/bs.coac.2017.02.001

[21] C. Sales, T. Portolés, L.G. Johnsen, M. Danielsen, J. Beltran, Olive oil quality classification and measurement of its organoleptic attributes by untargeted GC–MS and multivariate statistical-based approach, Food Chem., 271 (2019), pp. 488-496, 10.1016/j.foodchem.2018.07.200

[22] A. Martín-Sánchez, G. Valera-Marín, A. Hernández-Martínez, E. Lanuza, F. Martínez-García, C. Agustín-Pavón, Wired for motherhood: induction of maternal care but not maternal aggression in virgin female CD1 mice, Front. Behav. Neurosci., 9 (2015), 10.3389/fnbeh.2015.00197

[23] J. Beltrán, E. Serrano, F.J. López, A. Peruga, M. Valcarcel, S. Rosello, J. Beltran, E. Serrano, F.J. López, A. Peruga, M. Valcarcel, S. Rosello, Comparison of two quantitative GC–MS methods for analysis of tomato aroma based on purge-and-trap and on solid-phase microextraction, Anal. Bioanal. Chem., 385 (2006), pp. 1255-1264, 10.1007/s00216-006-0410-9

[24] L.G. Johnsen, J.M. Amigo, T. Skov, R. Bro, Automated resolution of overlapping peaks in chromatographic data, J. Chemom., 28 (2014), pp. 71-82, 10.1002/cem.2575

[25] L.G. Johnsen, P.B. Skou, B. Khakimov, R. Bro, Gas chromatography – mass spectrometry data processing made easy, J. Chromatogr. A, 1503 (2017), pp. 57-64, 10.1016/j.chroma.2017.04.052

[26] L. Lacalle-Bergeron, T. Portolés, C. Sales, M. Carmen Corell, F. Domínguez, J. Beltrán, J. Vicente Sancho, F. Hernández, Gas chromatography-mass spectrometry based untargeted volatolomics for smoked seafood classification, Food Res. Int., 137 (2020), p. 109698, 10.1016/j.foodres.2020.109698

[27] R.A. Harshman, PARAFAC2: mathematical and technical notes, UCLA Work, Pap. Phonetics., 22 (1972), pp. 30-44, http://www.bibsonomy.org/bibtex/2a964ff885ba59d4c7be518a3914f737a/threemode

[28] R.A. Harshman, M.E. Lundy, PARAFAC: parallel factor analysis, Comput. Stat. Data Anal., 18 (1994), pp. 39-72, 10.1016/0167-9473(94)90132-5

[29] B. Khakimov, J.M. Amigo, S. Bak, S.B. Engelsen, Plant metabolomics: resolution and quantification of elusive peaks in liquid chromatography–mass spectrometry profiles of complex plant extracts using multi-way decomposition methods, J. Chromatogr. A, 1266 (2012), pp. 84-94, 10.1016/j.chroma.2012.10.023

[30] V. Brust, P.M. Schindler, L. Lewejohann, Lifetime development of behavioural phenotype in the house mouse (*Mus musculus*), Front. Zool., 12 (2015), p. S17, 10.1186/1742-9994-12-S1-S17

[31] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, T.W.M. Fan, O. Fiehn, R. Goodacre, J.L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A.N. Lane, J.C. Lindon, P. Marriott, A.W. Nicholls, M.D. Reily, J.J. Thaden, M.R. Viant, Proposed minimum reporting standards for chemical analysis, Metabolomics, 3 (2007), pp. 211-221, 10.1007/s11306-007-0082-2

[32] H. Salais-López, M. Abellan-Alvaro, M. Bellés, E. Lanuza, C. Agustin-Pavon, F. Martínez-García, Maternal motivation: exploring the roles of prolactin and pup stimuli, Neuroendocrinology (2020), 10.1159/000510038

[33] J.R. Kelsey, N.H.P. Cnubben, J.J.P. Bogaards, R.B.H. Braakman, L.L.P. van Stee, K. Smet, The urinary metabolic profile of diethylene glycol methyl ether and triethylene glycol methyl ether in Sprague-Dawley rats and the role of the metabolite methoxyacetic acid in their toxicity, Regul. Toxicol. Pharmacol., 110 (2020), p. 104512, 10.1016/j.yrtph.2019.104512

[34] S. Lebreton, F. Borrero-Echeverry, F. Gonzalez, M. Solum, E.A. Wallin, E. Hedenström, B.S. Hansson, A.-L. Gustavsson, M. Bengtsson, G. Birgersson, W.B. Walker, H.K.M. Dweck, P.G. Becher, P. Witzgall, A Drosophila female pheromone elicits species-specific long-range attraction via an olfactory channel with dual specificity for sex and food, BMC Biol., 15 (2017), p. 88, 10.1186/s12915-017-0427-x

[35] J.R. Aldrich, M.C. Rosi, F. Bin, Behavioral correlates for minor volatile compounds from stink bugs (*Heteroptera: pentatomidae*), J. Chem. Ecol., 21 (1995), pp. 1907-1920, 10.1007/BF02033851

[36] R.A. Hayes, B.J. Richardson, S.G. Wyllie, To fix or not to fix: the role of 2-phenoxythanol in rabbit, *Oryctolagus cuniculus*, chin gland secretion, J. Chem. Ecol., 29 (2003), pp. 1051-1064, 10.1023/A:1023836319677

[37] M. Sam, S. Vora, B. Malnic, W. Ma, M.V. Novotny, L.B. Buck, Odorants may arouse instinctive behaviours, Nature, 412 (2001), 10.1038/35084137

[38] P. Mitra, S. Das, R. Debnath, S.H. Mobarak, A. Barik, Identification of *Lathyrus sativus* plant volatiles causing behavioral preference of *Aphis craccivora*, Pest Manag. Sci., 77 (2021), pp. 285-299, 10.1002/ps.6018

[39] X.-Y. Zhang, X.-Q. Zhu, S.-H. Gu, Y.-L. Zhou, S.-Y. Wang, Y.-J. Zhang, Y.-Y. Guo, Silencing of odorant binding protein gene AlinOBP4 by RNAi induces declining electrophysiological responses of *Adelphocoris lineolatus* to six semiochemicals, Insect Sci., 24 (2017), pp. 789-797, 10.1111/1744-7917.12365

Supplementary Information

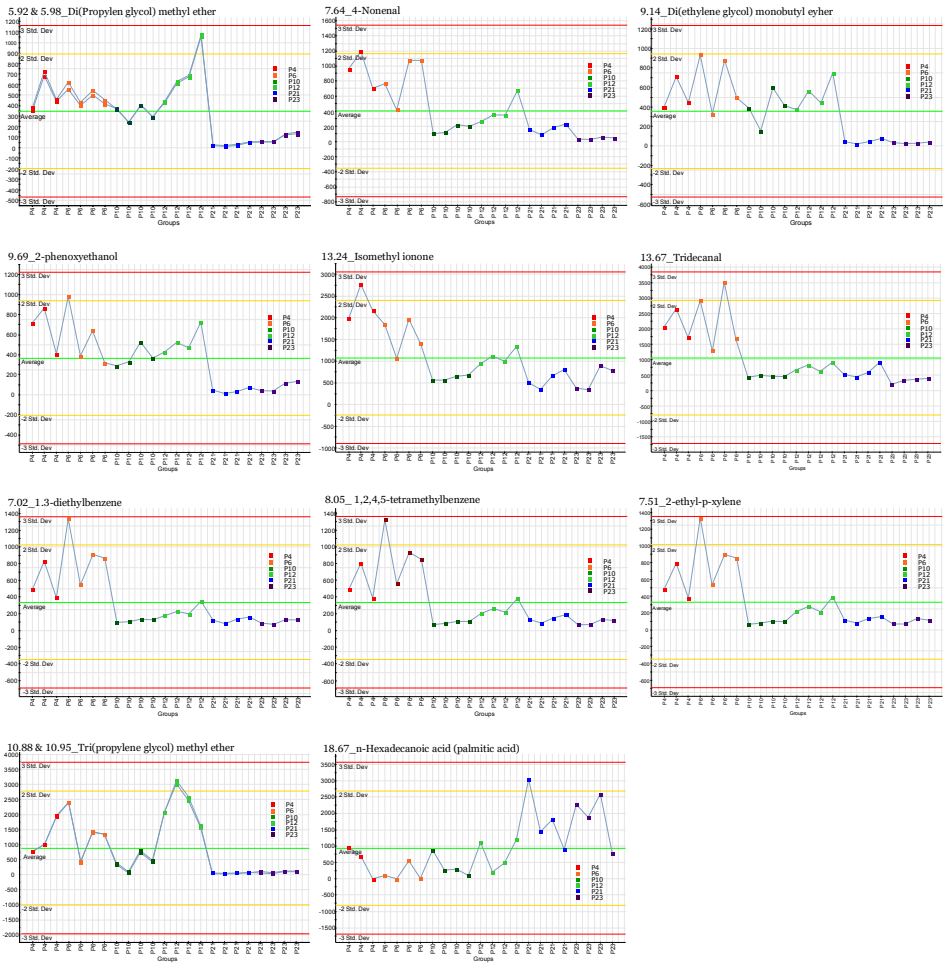


Fig. S3. *Variable trend plot of the 13 identified markers corresponding to 11 compounds from the mice pup volatolome extraction (areas obtained by PARAFAC and corrected with the internal standard and mice pup the total weight).*

Fig. S4. *Comparison of areas between pups volatolome extraction and soiled-cotton litter extraction of the first two weeks, for the 13 identified markers corresponding to 11 compounds separated. In orange shades the average area obtained from the samples of the same week of direct pups volatolome extraction and green shades from the soiled-cotton litter extraction (both with areas corrected by internal standard and mice pup the total weight corrected).*

5.3 Scientific article VI

Talanta (2021) Submitted

Untargeted metabolomics approach to putative pheromones in mice. Part 2: using UHPLC-IMS-QTOF MS for surface body samples to identify low-volatility chemosignals

Leticia Lacalle-Bergeron[a], Rafael Goterris-Cerisuelo[b], Joaquin Beltran[a], Juan V Sancho[a], Cinta Navarro-Moreno[b], Fernando Martinez-Garcia[b], Tania Portolés[a]*

[a] Enviromental and Public Health Analytical Chemistry, Research Institute for Pesticides and Water (IUPA), University Jaume I, Av. Sos Baynat S/N, 12071 Castellón de la Plana, Spain

[b] Laboratory of Functional Neuroanatomy (Unitat Mixta NeuroFun-UV-UJI), Predepartamental Unit of Medicine, Universitat Jaume I, Av. Sos Baynat S/N, 12071, Castellón de la Plana, Spain

Abstract

The present study is focused in the determination of low-volatile chemosignals excreted or secreted by mouse pups in their early days of life involved in maternal care induction in mice adult females. Untargeted metabolomics was employed to differentiate between samples collected with swabs from facial and anogenital area from neonatal mouse pups receiving maternal care (first two weeks of life) and the elder mouse pups in the weaning period (4th week old). The sample extracts were analysed by liquid chromatography (LC) coupled to ion mobility separation (IMS) in combination with high resolution mass spectrometry (HRMS). After data processing with Progenesis QI and multivariate statistical analysis, five

markers present in the first two weeks of mouse pups life and putatively involved in materno-filial chemical communication were tentatively identified: arginine, urocanic acid, erythro-sphingosine (d17:1), sphingosine (d18:1) and sphinganine. The four-dimensional data and the tools associated to the additional structural descriptor obtained by IMS separation were in great help in the compound identification. The results demonstrated the great potential of UHPLC-IMS-HRMS based untargeted metabolomics to identity putative pheromones in mammals.

Keywords

untargeted metabolomics, ion mobility, chemical signalling, mouse pheromones, maternal care, HRMS

1. Introduction

The role of chemicals for intra- and inter- specie communication has been widely demonstrated through the animal kingdom. Many species use true pheromones [1], e.g. chemicals excreted and/or secreted that elicit stereotyped behavioural and/or neuroendocrine-developmental responses in conspecifics [2]. Macrosmatic mammals such as rodents, possess two nasal chemosensory systems for monitoring of chemicals in their environment, including pheromones: the olfactory epithelium (OE), which detects a myriad of volatiles present in the air; and the vomeronasal organ (VNO), which detects compounds with high and low volatility [3]. Screening the response of the VNO of mice to different chemicals revealed that vomeronasal cells also respond to many volatile pheromones as well as to small molecules secreted by predators (carnivores) [4]. The last thirty years of investigation have uncovered many molecules with different physico-chemical properties mediating a wide variety of social interactions including as kin and individual recognition, sexual attraction, dominancy, aggression among others [5]. In addition, several studies have demonstrated a role of chemosensing in the interaction of adult mice with pups. For example, some chemosignals apparently mediate pup killing by sexually-naïve males [6], which seems largely dependent on signalling in cells of the vomeronasal organ [7]. Moreover, induced anosmia results in dramatic changes in the response of females to pups resulting from maternal neglect to systematic pup killing [8,9]. However, it is surprising that no study has been carried out to unveil the chemical cues involved in materno-filial communication, in spite of the crucial role of maternal behaviour in pups survival and development [10].

In our first paper on that issue, it was explored volatile putative pheromones involved in materno-filial communication in mice [11] by comparing the volatolome of neonatal pups, receiving dedicated care by their mothers, with 4-week pups, e.g. the age of weaning. In this study, the aim was focused on those molecules with lower volatility present in the body surface of neonatal pups (anogenital and facial regions) that could induce maternal care in mice, and that they decrease with age until weaning. To do so, untargeted metabolomic approach was used for screening metabolites, which can act as a pheromone. This comprehensive methodology aims to

differentiate between samples coming from an organism that has suffered changes in their metabolite fingerprinting in response to internal (genetics, disease, growth…) or external (environmental, diet…) perturbations [12,13]. As our objective was to focus on low-volatile compounds excreted or secreted by mouse pups, reversed phase liquid chromatography (RP-LC) was the technique of choice for the separation of medium to non-polar compounds that can constitute some of the released pheromones. The evolution of untargeted metabolomics has been possible thanks to the implementation of powerful statistical tools such as multivariate statistical analysis [14], as well as the great improvement in analytical instrumentation [15]. In particular, ion mobility mass spectrometry (IMS) introduction to the conventional liquid chromatography coupled to high resolution mass spectrometry (LC-HRMS) has an emerging role in untargeted metabolomics in the generation of multidimension data to support metabolites identification [16]. UHPLC-IMS-HRMS combination provides an extra separation dimension through the drift time (DT), later converted into collisional cross section value (CCS, $Å^2$), to the conventional three-dimensional data formed by retention time (Rt), accurate mass (m/z) and intensity from LC-HRMS. This additional molecular descriptor is dependent on the individual size, shape and charge of a molecule, and therefore, it can be transversal between instruments regardless of the chromatographic technique employed [17]. In fact, the CCS value is being added to the compound databases together with the spectral information; and various CCS prediction tools have been developed to help in the compound identification process [18,19]. Therefore, in this work an untargeted metabolomic approach in combination with UHPLC-IMS-HRMS was applied to samples collected from facial and anogenital area of mouse pups, in order to identify putative chemosignals with low volatility involved in the maternal care induction, not covered in the first part of the study. In addition, the contributions of IMS and its associated tools to the identification process are explored.

2. Materials and methods

2.1. Chemicals and reagents

Methanol LC-MS grade was purchased from Scharlab (Barcelona, Spain), as well as formic acid eluent additive for LC-MS. To obtain HPLC-grade water a Milli-Q water purification system (Millipore Ltd., Bedford, MA, USA) was employed. Leucine-enkephaline HPLC-grade (mass axis calibration) was purchased from Sigma-Aldrich (Saint Louis, MO, USA).

2.2. Animals, experimental design and sampling

To perform the study, a total of n = 4 mouse female (CD1 strain, 10-week-old) (Janvier Labs, France) were employed. Animals were treated throughout according to the European Union Council Directive of June 3rd, 2010 (6106/1/10 REV1) and the study was approved by the Committee of Ethics on Animal Experimentation of the Jaume I University of Castellón, where the experiments were carried out and, ultimately, by the Valencian *Conselleria d'Agricultura Medi Ambient, Canvi Climàtic i Desenvolupament Rural* (code 2019/VSC/PEA/0049).

After mating with an adult male, the pregnancy was carefully controlled following the conditions extensively described in Lacalle-Bergeron et al., 2021 [11]. This second study was carried out simultaneously to the one referred in our previous article and used the same animals.

Two of the females delivered 17 days before the other two as planned, in order to process in parallel pups of different ages. In this way, the first day of experimentation, right after the volatolome extraction for our previous study [11], the surface of the body pheromones was sampled simultaneously in neonatal pups (4-day old) and pups at the age of weaning (21 days-old), by gently rubbing their anogenital region with a cotton swab for 1 minute, and then the orofacial region (eyes, nose, mouth) with another cotton swab. This sampling procedure was performed for four consecutive days per week, thus obtaining samples from pups of 4-to-7-days old (1st week pups) in parallel to 21-to-24-days old (4th week pups). The following week, the younger pups were sampled again, thus obtaining samples from 10-to-13-days old pups (2nd week pups).

2.3. Extraction of low-volatile putative pheromones

Three swabs coming from the same body area of 3 pups selected randomly from the same litter (siblings, with independence of the sex) were soaked together in 1500 µL of acetone and further extracted in a thermostatic water bath at 25 ºC with ultrasound assistance for 20 min (**Figure 1**).Then, the organic solvent was evaporated to dryness at 35 ºC under a gentle stream of nitrogen, to finally reconstitute the residue in 150 µL of methanol for instrumental analysis. In total, four samples per day and per area were obtained, and 4 days were studied per week (**Table 1**). Each day of analysis, blank extracts were performed following the same procedure with three clean swabs. From each sample extract, 50 µL were aliquoted to be pooled as a quality control sample (QC), which was injected at the beginning of the sample batch for chromatographic column conditioning and every 10 samples to ensure the stability and repeatability of the system.

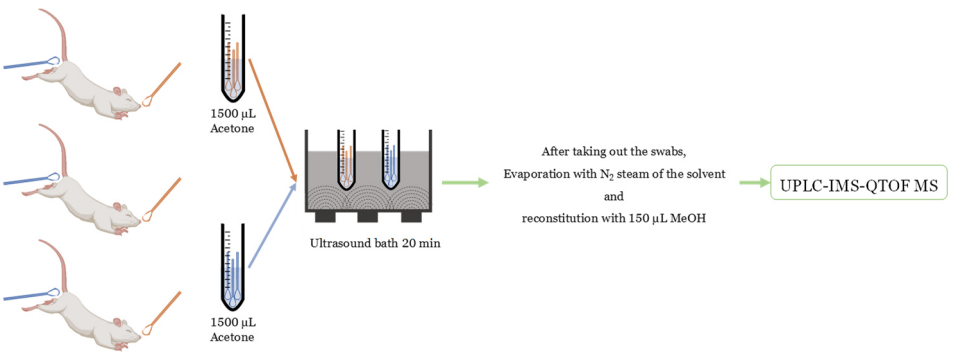


Figure. 1. *Experimental design used for alive mice pup low-volatility putative pheromones collection. Swabs gently rubbed in the facial area are coloured in orange and those for the anogenital area in blue. Three swabs from different mice pups but from the same zone (facial or anogenital) were extracted together.*

2.4. Instrumentation

Ultra-high performance liquid chromatography (UHPLC) with ACQUITY UHPLC I-Class system (Waters, Milford, MA, USA) was coupled to a VION® IMS QTof (Waters, Manchester, UK), ion mobility hybrid quadrupole time-of-flight (IMS-QTOF) high resolution mass spectrometer (UHPLC-IMS-HRMS) using electrospray interface operating in positive

ionization mode (ESI+). Equipment control, data acquisition and processing were performed using UNIFI software (V.1.9.2, Waters, Manchester, UK).

The chromatographic separation employed was reversed phase liquid chromatography (RPLC) using a CORTECS® C18 fused-core 2.7 µm particle size analytical column 100 x 2.1 mm (Waters). The flow rate was set at 0.3 mL/min, column oven at 40 ºC and 1 µL sample injection volume was selected. The mobile phases employed were A = H_2O and B = methanol, both with 0.01 % formic acid. The gradient changed from 10 % B at t = 0 min to 90 % B at t = 14 min, holding it for 2 min, and at t = 16.01 min B decreased to 10 % and held for 2 min; total run time 18 min.

Regarding the HRMS analyser, the capillary voltage was set at 0.7 kV in ESI+ ionization mode. Nitrogen was used as desolvation gas, nebulizing gas, mobility gas and collision gas. Source temperature was set to 120 ºC and desolvation gas to 550 ºC with a flow rate of 1000 L/h. The HRMS analyser operated in MS^E acquisition mode combined with ion mobility separation ($HDMS^E$). As normal MS^E acquisition mode, it is a type of data independent acquisition (DIA) analysis where two functions are acquired sequentially: low energy function (LE), with a fixed collision energy at 6 eV, and high energy function (HE) with a ramp of collision energy from 28 to 56 eV. Both acquisitions were performed from 50 to 1000 Da and at 0.3 s scan time.

"Major Mix IMS/T of Calibration kit" (Waters) infused at 20 µL/min flow rate was used for calibration of mass axis and drift time. For automated accurate mass correction, 100 ppb Leucine-Enkephalin acetonitrile:H_2O (50:50, *v/v*) at 0.01 % formic acid solution was infused at 20 µL/min into the system through the lock-spray needle and measured every 5 min with 0.3 s scan time during the sample injection (ensuring a measurement at the beginning and at the end of each run), monitoring the protonated molecule *m/z* 556.27658.

2.5. Data processing and statistical analysis

Raw data from VION instrument (*.uep*, UNIFI, Waters) were imported into Progenesis QI (v3.0, NonLinear Dynamics, Waters, UK) for data processing. Progenesis QI software automatically performs 4D peak picking (based on *m/z*, retention time, drift time and intensity); retention time

alignment (with QC samples as references, except the first 9 QC injection employed for column stabilization purposes); and response normalisation. Peak picking conditions were set as follows: all runs, sensitivity automatic (level 3), minimum chromatographic peak width 0.1 min and retention time limits from 0.4 to 17 min from the total run time of 18 min. The deconvolution was applied attending to the selected adducts and ion forms for positive ESI+: $[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, $[M-H_2O+H]^+$, $[2M+H]^+$ and $[2M+Na]^+$. The samples were divided in groups in the "Experimental Design Setup" section of the software according to the experiment and statistical comparison, starting by a basic classification depending on the area (facial or anogenital) and the week of life (week 1, week 2 or week 4), following the "Between -subject Design" comparison (samples from a given subject appear in only one condition). One-way ANOVA calculation followed by a false discovery rate (FDR) was performed to test differences among the groups, setting the level of statistical significance at 95 % (*p-value* <0.05).

The main work of statistical analysis was performed with EZinfo (v3.03, Umetrics, Sweden), focusing on multivariate statistical analysis. After pareto-scaling of the data, the process started by unsupervised Principal Component Analysis (PCA) in order to check the correct acquisition of the samples based on the grouping of the QC replicates in the centre of the plot; and detect possible outliers. Then, Partial Least Square-Discriminant Analysis (PLS-DA) to maximize the separation between the pre-determined groups and the validation of the model was performed by leaving-1/7-out cross-validation approach. Finally, Orthogonal PLS-DA (OPLS-DA) was carried out to highlight the most significant markers between the faced groups (threshold *p(corr)* ≥ |0.6| and *p[1] loading* ≥ |0.1|).

3. Results and discussion

3.1. Experimental set up for low volatile putative pheromone obtention

As explained in the previous research article [11], the procedure employed did not interfere with pup's growth and comfort (pups gained weight according to the standards) (**Table 1**). The sampling from pups was carried out quickly and accurately by the personnel accredited in animal

handling so that the process was as stress-free as possible for the animals under study. The pups were returned after the sampling to its mother and kept under the appropriate and controlled conditions for its development until the next day of sampling.

Table 1. M*ice pup characteristics for each experiment day*

Experiment	Nº of replicates	Nº of mice pups per replicate	Postnatal week	Postnatal day	Mean mice pup weight
1	4	3[a]	1	4	2.6 ± 0.4 g (*n* = 24)
2	4	3[a]	1	5	3.0 ± 0.4 g (*n* = 24)
3	4	3[a]	1	6	3.4 ± 0.5 g (*n* = 24)
4	4	3[a]	1	7	3.9 ± 0.5 g (*n* = 24)
5	4	3[b]	2	10	5.2 ± 0.8 g (*n* = 18)
6	4	3[b]	2	11	5.6 ± 0.7 g (*n* = 18)
7	4	3[b]	2	12	5.9 ± 0.7 g (*n* = 18)
8	4	3[b]	2	13	6.2 ± 0.8 g (*n* = 18)
9	4	3[c]	4	21	12.5 ± 0.8 g (*n* = 9)
10	4	3[c]	4	22	13.3 ± 0.8 g (*n* = 9)
11	4	3[c]	4	23	14.5 ± 0.8 g (*n* = 9)
12	4	3[c]	4	24	15.6 ± 0.9 g (*n* = 9)

[a] *3 mice selected aleatorily from the 8 mice pups of each replicate for the volatile extraction explained in Lacalle-Bergeron et al. 2021*
[b] *3 mice selected aleatorily from the 6 mice pups of each replicate for the volatile extraction explained in Lacalle-Bergeron et al. 2021*
[c] *3 mice selected aleatorily selected for the volatile extraction explained in Lacalle-Bergeron et al. 2021 where the same for this second experiment*

Based on the premise that some of the compounds with a putative role as pheromones inducing maternal care could not be very volatile, this second experiment was focused on low volatile compounds of exocrine glands secreting an aqueous medium, such as saliva or sweat, or excreted in urine. In both cases, these compounds are expected to be rather polar compounds. Less polar compounds, which could be significant for chemical communication, could also be secreted by sebaceous glands.

The starting hypothesis was that mouse pup pheromones are secreted/excreted during the first days of life, when dams exhibit dedicated pup care, but their release decreases with age until weaning.

Besides the molecules that are excreted in urine/faeces, most chemosignals must derive from secretions of exocrine glands: salivary, lacrimal and other glands located in the chin perineal and genital areas. This is the reason why sampling was focused on the anogenital and facial areas. To extract them from the skin, swabs were rubbed around each area during 1 min and samples coming from each of them were treated separately. Afterwards, the extraction of the collected molecules with the swabs was carried out in acetone. The choice of the solvent was based on its compatibility with molecules in medium range of polarity and the ruling out of those too large to be covered in a metabolomic study, such as proteins.

As the pheromones should be released in a very similar way in all pups of the same age, and regardless the sex, 3 mouse pups of the same day of age were randomly chosen for each sample extraction. The collected sample from each area of those 3 pups were extracted in the same 1500 µL of acetone to increase the concentration (**Figure 1**).

RPLC with a C18 column was selected due to well-known behaviour, good robustness and its ability to cover a wide range of compounds. Due to the high sensitivity of the instrument employed and the small injection volume required (1 µL), no additional sample treatment steps were necessary, such as evaporation and reconstitution in more compatible solvent for the chromatography. As it is a common practice in chromatography-MS-based metabolomics, the quality of the results and the analytical process were monitored by means of injection of mechanic replicates (every 10 samples) of quality control sample (QC sample, a homogeneous pool of all treated samples) during the chromatographic run. Moreover, it is highly recommended to perform the sample treatment and the acquisition in a randomised way. Although this could not be possible in the sample treatment, as the extraction was performed as the mouse pups grew up, it was applied for the UHPLC-IMS-HRMS acquisition of the data, assuring that the possible instrumental drift during the sequence did not affect to a group more than others and minimizing the bias.

3.2. Data processing and statistical analysis

The data set acquired with UNIFI software (Waters, UK) was exported to **.uep* format (UNIFI export package). To the best of our knowledge, only Progenesis QI (nonlinear Dynamics, UK) program is able to process and interpret four-dimensional data (Rt, *m/z*, intensity and DT) for *-omics* purposes. After the data import, Progenesis QI performs 4D peak picking, followed by retention time alignment and finally normalisation. For the retention time alignment one sample between the QC replicates was selected as reference (except the first 9 QCs samples injected at the beginning of the chromatographic run for column stabilization). All samples had a high and very good alignment (all above 90 %) and normalisation method for the aligned data was "normalisation to all compounds". The resulting data matrix consisted in the detection of 6094 signals that Progenesis QI assigned to a total of 5454 features, due to the automatic deconvolution performed with the specified adducts selected in the setting of the peak picking. Therefore, for those features where more than one adduct was detected the program will annotate them as *xx.xx_yyy.yyyy*n (*xx.xx* being the retention time in minutes and *yyy.yyyy* the exact neutral mass calculated when more than one adduct ion is found for the same compound); and for single ions *xx.xx_zzz.zzzzzm/z*, *zzz.zzzz* being the accurate mass. The 5454 data set was reduced to 3647 features after the removal of all features with poor stability in the QCs samples, keeping all features with a relative standard deviation (% RDS) below the 30 % within the QCs.

An exploratory visualization of the data was performed by unsupervised multivariate analysis PCA in order to observe trends, inherent grouping of the data and possible outliers. As QC is considered as an "average" sample (constituted by a pool of all analysed samples), and it is injected after every 10 samples (n = 12 replicates), the clustering in the centre of the PCA score demonstrated the correct acquisition of the run, determining that the inherent differentiation between the samples does not come from an instrumental drift during the acquisition of the data. QC samples grouping near to the centre can be observed in **Figure 2**, PCA score of component 1 vs component 2, proving the proper performance of the analytical system. Nevertheless, no inherent differentiation was observed between the groups of

samples. Only a slight separation of blank samples can be noticed in the upper left quadrant. Twelve principal components were necessary to explain the 88 % of the total variance. The blanks data were removed from the subsequent statistical analysis, although they were consulted later to validate the features highlighted in the statistics as markers by checking their absence in these samples, demonstrating that they did not come from the swabs employed or any other factor during the experimental procedure.

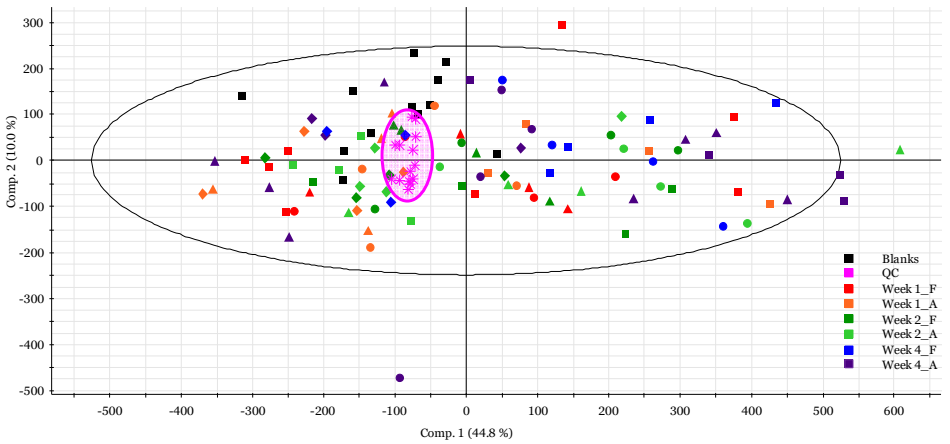


Figure. 2. *PCA score plot component 1 vs component 2 of all analysed samples from anogenital and facial areas. The QC samples (QC ■) are grouped and centred in the plot. The samples were coloured by week and in the same tones by area: first week facial area (Week 1_F■), first week anogenital area (Week 1_A■), second week facial area (Week 2_F■), second week anogenital area (Week 2_A■), fourth week facial area (Week 4_F■) and fourth week facial area (Week 4_A■). Blank samples were also present in the PCA (Blank ■). The variance explained was 88 %.*

Then the statistical analysis was focused on the differentiation of the samples by the age of mice pup, analysing samples coming from facial and anogenital separately. Taking into account the information of the volatolome separation from the first part of the study, univariate statistical analysis was applied obtaining 338 and 912 statistically significant features for the differentiation by week of the samples in the facial area and in the anogenital area, respectively. From these reduced data matrices, a second PCA was applied to interrogate the data (**Figure 3A** and **B**). Four components were enough to explain the 83 % of the variance for the differentiation by the

compounds collected from the facial area and eight components for the 87 % of the variance explained for the anogenital area. Colouring the samples according to the week of life in these two PCA, a separation by age can be observed, the samples coming from the aged pups (week 4) are clearly differentiated from the samples coming from the younger ones. A slight differentiation of samples from the first and second week of life might be appreciated in the samples coming from the facial area by the 2nd component (**Figure 3A**) that it is not observed for the anogenital region (**Figure 3B**). Some samples were taken as possible outliers, especially in the PCA from anogenital samples, and they were studied in the following analysis before considering eliminating any of them from the statistics.

Subsequently, PLS-DA was applied to both data sets, attempting to target a grouping by week of life in this supervised multivariate statistical analysis. The PLS-DA score plots facing the first two latent variables are shown in **Figure 4A** and **4B**, where classification of pups by weeks is appreciated. For PLS-DA based on the compounds extracted from the facial region it was obtained a 64 % of the variance explained and a 52 % of the variance predicted with three latent variables. The PLS-DA from anogenital area obtained a variance explained of 80 % and a predicted of 58 % with five latent variables. Although these results in the variance might seem lower than desirable and close to the accepted limit of 50 % for variance explained and 40 % for predicted for biological models [20], this is mainly due to the difficulty in separating the groups of the first two weeks. However, a clear differentiation of the groups of samples that come from the oldest pups can be observed along the first latent variable of both PLS-DAs, the one that explains the greater differentiation of the model (52.3 % for facial area and 36.2 % in anogenital area).

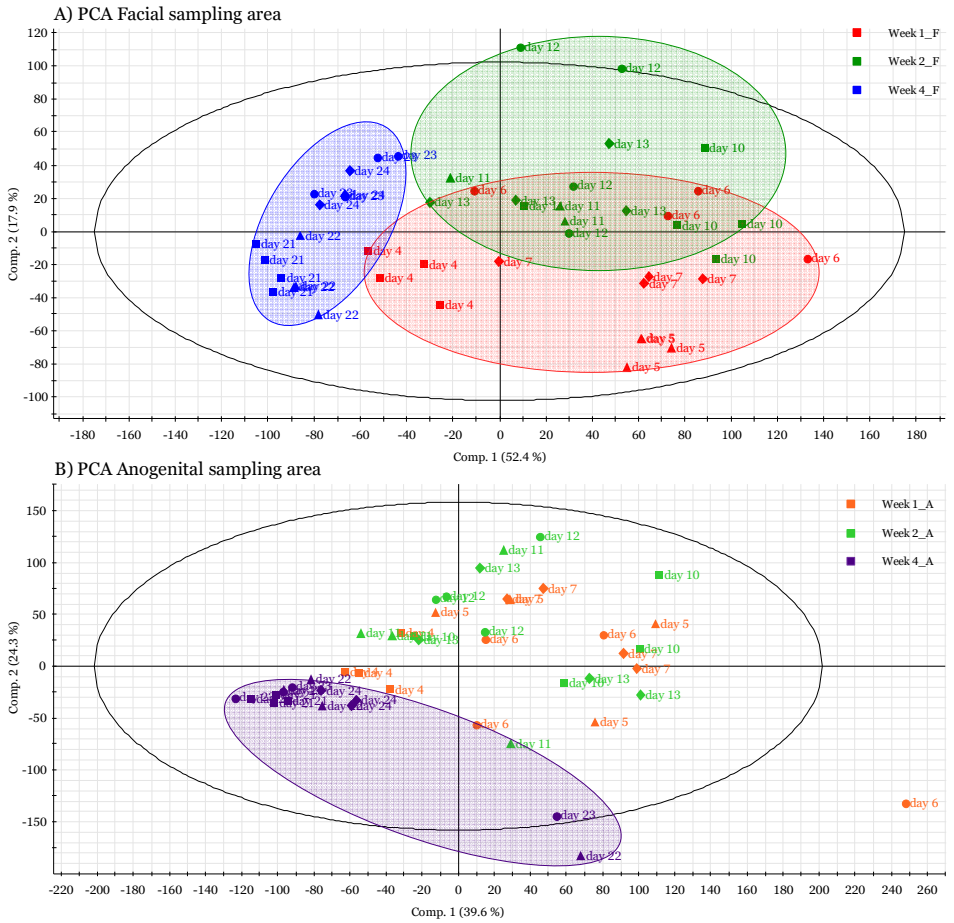


Figure. 3. *PCA score plot component 1 vs component 2 of from A) Facial and B) Anogenital areas coloured by week: 1st week facial area (Week 1_F■), 2nd week facial area (Week 2_F■), and 4th week facial area (Week 4_F■); and for anogenital area 1st week (Week 1_A■), 2nd week (Week 2_A■) and 4th week (Week 4_A■).The shape of the marker changed according to the day of the week (1st day of the week □, 2nd day Δ, 3rd ○ and 4th day ◇). The variance explained was 83 % and 81 % for facial and anogenital area respectively.*

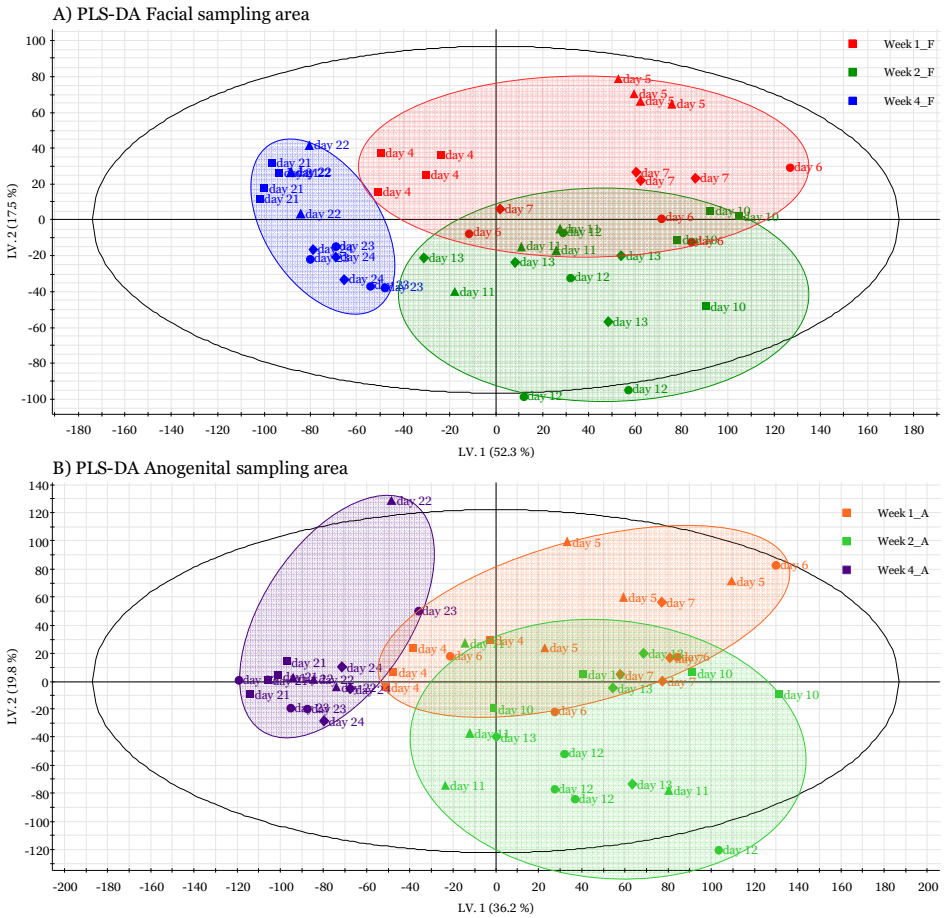


Figure. 4. *PLS-DA score plot latent variable 1 vs latent variable 2 of from A) Facial and B) Anogenital areas coloured by week: 1st week facial area (Week 1_F■), 2nd week facial area (Week 2_F■), and 4th week facial area (Week 4_F■); and for anogenital area 1st week (Week 1_A■), 2nd week (Week 2_A■) and 4th week (Week 4_A■).The shape of the marker changed according to the day of the week (1st day of the week □, 2nd day Δ, 3rd ○ and 4th day ◇). Being R^2 is coefficient for variance explained and Q^2 is coefficient for variance predicted, the following results were obtained: R^2 = 64 % and Q^2 = 49 % for facial sampling area and R^2 = 80 % and Q^2 = 58 % for anogenital sampling area.*

Following the objective on finding molecules with higher presence in the first weeks, a binary differentiation was attempted with OPLS-DA, starting by week 1 vs week 4 expecting that this facing would be the most significant; and followed by week 2 vs week 4 differentiation, obtaining in all of them a variance explained above the 80 % and above the 70 % for the variance predicted (**Table 2**, **Figure 5** and **6**). The S-plots from the OPLS-DA allowed an easiest selection of putative markers in each binary differentiation. The most differentiating features are those with a *p(corr)* closer to 1 or -1 (y-axis of the S-Plot), and with higher intensity the ones with a *p [1] loading* closer to 1 or -1 (x-axis of the S-Plot). All features with a *p(corr)* higher than |0.6| and a p [1] loading above |0.1|were selected as putative markers (**Figure 5** and **6**). A total of 17, 14, 12 and 16 features were highlighted in the corresponding S-Plots of facial sampling area week 1 vs week 4 and week 2 vs week 4 and anogenital sampling area week 1 vs week 4 and week 2 vs week 4, respectively. In sum, a total of 34 features highlighted as putative markers, since the features with higher abundance for week 1 used to be the same as the ones more abundant in week 2, hence the little differentiation in the PLS-DA of these two groups. In fact, the 1st week vs 2nd week facing in the OPLS-DA was also applied with poor statistical results. Since the objective of the study was not to confront these two groups, no further investigation was made of said separation.

Table 2. *Parameters of OPLS-DA models.*

Statistical model / Characteristics	***OPLS-DA model diagnostics***			
	Facial sampling area		Anogenital sampling area	
	Week 1 vs Week 4	Week 2 vs Week 4	Week 1 vs Week 4	Week 2 vs Week 4
	n=16 vs 16	n=16 vs 16	n=16 vs 16	n=16 vs 16
Components	2	3	4	4
Goodness-of-fit parameter - R^2X	69 %	74 %	74 %	77 %
Variance explained – R^2Y (cum)	80 %	92 %	94 %	96 %
Variance predicted – Q^2 (cum)	71 %	86 %	83 %	88 %

R^2 - fit how model fits the data and Q2 – predictive ability, by seven-round internal cross-validation as default option of EZinfo software (Umetrics, Sweden)

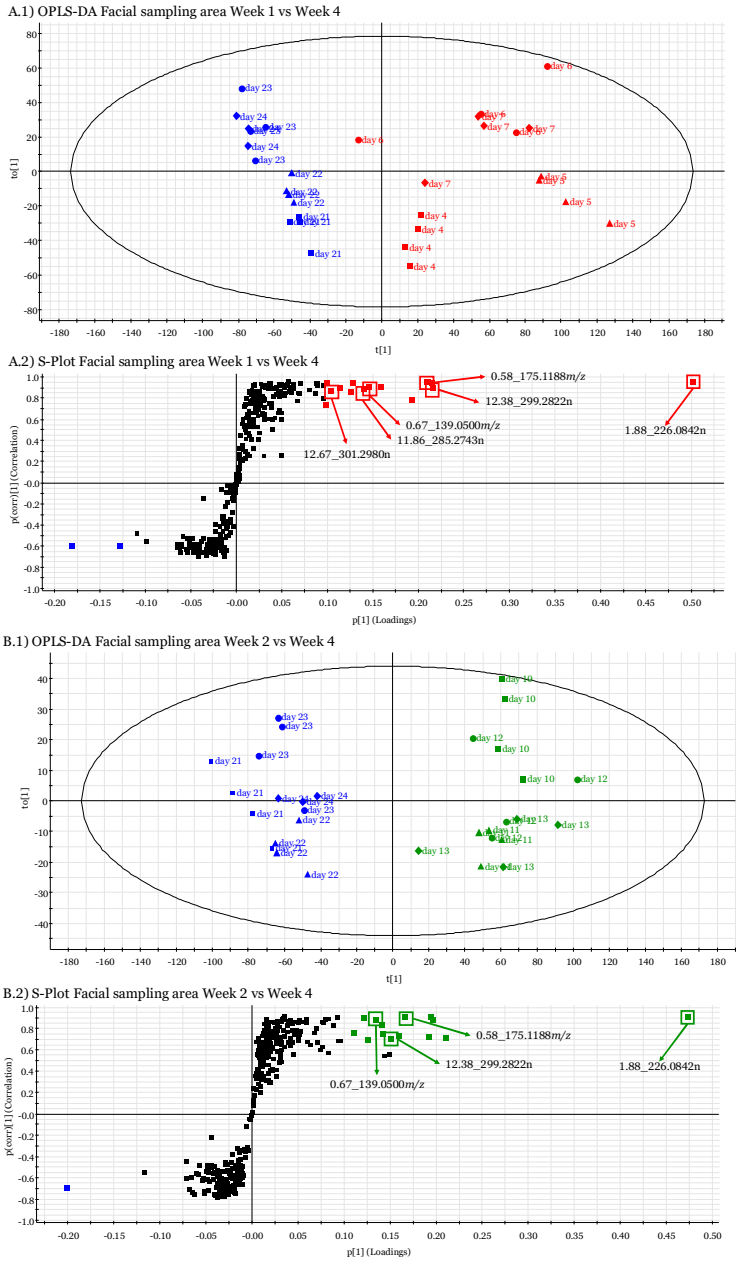


Figure. 5. *OPLS-DA score plots and S-Plots from A) Facial sampling area Week 1 vs Week4 and B) Week 1 vs Week4.Coloured by week: 1st week (Week 1_F■), 2nd week (Week 2_F■), and 4th week (Week 4_F■).The shape of the marker in the OPLS-DA changed according to the day of the week (1st day of the week □, 2nd day Δ, 3rd ○ and 4th day ◇). Being R^2 and Q2 coefficients are founded in* ***Table 2****. The coloured features in the S-plots were the ones selected as putative markers, their colours depend on the week where they are highlighted as markers (see above).*

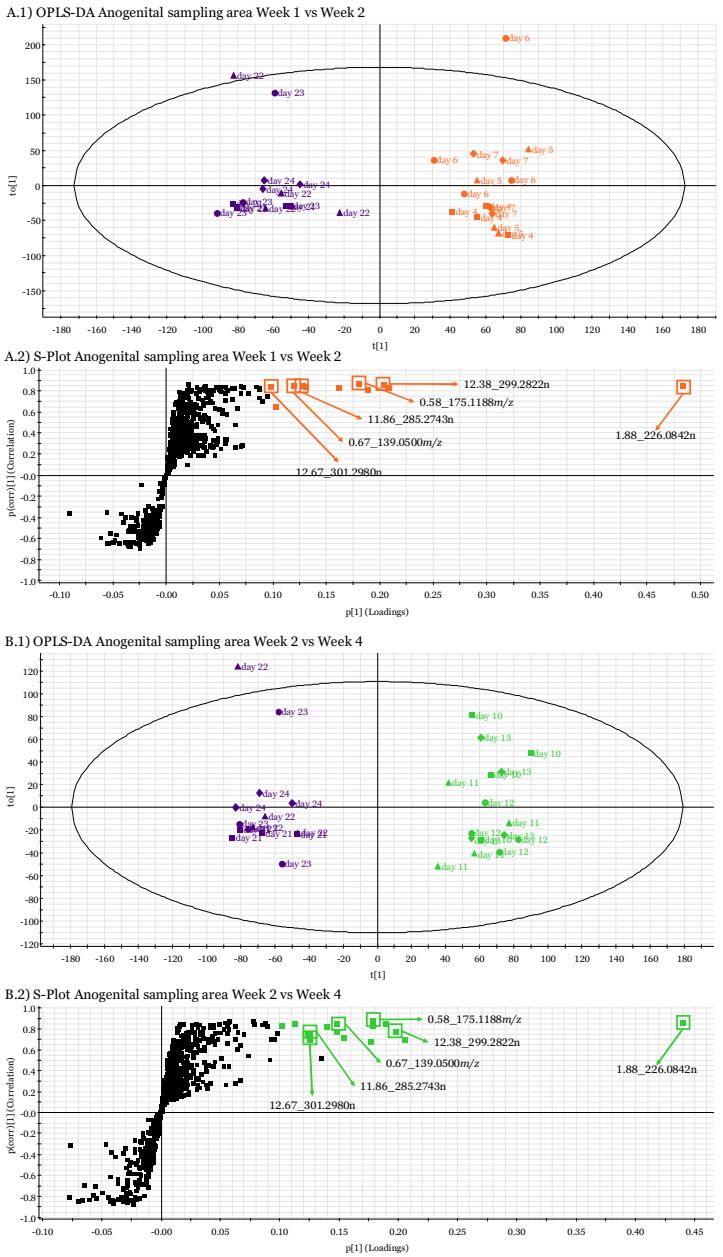


Figure. 6. *OPLS-DA score plots and S-Plots from A) Anogeniatal sampling area Week 1 vs Week4 and B) Week 1 vs Week4.Coloured by week: 1st week (Week 1_A■), 2nd week (Week 2_A■), and 4th week (Week 4_A■).The shape of the marker in the OPLS-DA changed according to the day of the week (1st day of the week □, 2nd day Δ, 3rd ○ and 4th day ◇). Being R^2 and Q2 coefficients are founded in* ***Table 2****. The coloured features in the S-plots were the ones selected as putative markers, their colours depend on the week where they are highlighted as markers (see above).*

Moreover, most of the highlighted markers in the facial sampling area were the same as in anogenital area. One reason could be the presence of similar glands in both regions, as sweat and sebaceous glands. On the other hand, the continuous licking of pups by dams could cause that the compound released by one of the areas ended up in the other one. Before starting with the elucidation process, it was studied the presence of these features in the blank samples, thus discarding 23 out of 34 markers since their signals in these samples were relevant enough to question their validity as markers. All remaining features were markers of the 1st and the 2nd week when they were faced to the 4th week. Nevertheless, the list can be further reduced as some type of adducts were not set in the peak picking processing, and therefore not clustered as a single compound; or because of the presence of some in-source fragments (formed prior to the IMS, and therefore with different DT) detected as a different feature.

3.3. Elucidation Process

The remaining 19 putative markers were studied in order to obtain a tentative identification. To do so, features were annotated with Progenesis QI identification tool based on the deconvoluted $HDMS^E$ spectra and CCS data by search and comparison with exact mass, MS/MS and CCS (if available) information in compound and spectral databases. Moreover, the experimental information about the parent ion, fragmentation spectra and CCS were carefully reviewed from UNIFI raw data. The identification level of each marker (**Table 3**) was set according the criteria of our laboratory based on Celma et al. (2020) classification [21]. In **Table 3** the experimental and statistical data for each tentatively identified marker are recorded.

Markers 1 and 2 received an identification level 2a as Urocanic acid and Arginine respectively, by match MS/MS and CCS data with available libraries (constituted of experimental data for fragmentation and CCS for both compounds). These two markers had a *p(corr)* higher than 0.85 for week 1 and 2 in both facial sampling area and anogenital (see **Table 3**). As an example, the elucidation process for marker 1 is explained hereafter. Progenesis QI performed a deconvolution of the spectra from LE (6 eV collision energy) and HE (ramp of collision energy from 28 to 56 eV) and compared the results and CCS data with the available compound/spectral

databases. Even the best annotation provided by Progenesis QI (defined by the identification score of each candidate) was carefully reviewed with raw data. **Figure 7** shows the obtained UNIFI raw data for Marker 1 (0.67_139.0500*m/z*, CCS 131.15 Å^2), including the extracted ion chromatograms and HDMS^E spectra. As IMS takes place prior to the collision cell, the product ions from fragmentation are linked to their parent ion by the DT of the last one. For this reason, it is possible to obtain cleaner spectra without the interference in the LE spectra of co-eluting ions and their fragments in the HE spectra. Figure 6A and 6B shows the LE and HE spectra not only filtered by the retention time (0.67 ± 0.04 min), as it is usual in conventional LC-HRMS, but also by the DT (3.09 ± 0.21 ms), obtaining cleaner spectra for both energy functions. As part of the verification, the extracted ion chromatograms of both fragments were obtained from the HE function (**Figure 7A**) to verify that the peak shape is similar to the one from parent ion. Both fragments are almost the only fragments found in the experimental spectra of databases. All this process allowed the selection of Urocanic acid as tentative identification among the other candidates (e.g. 4-nitroanuline or methyl-2-pyrimidine carboxylate, isomeric compounds with different fragmentation patterns). Moreover, there was a match with a delta error of +1.35 % (maximum tolerance 2.5 %) with the available experimental CCS library for Progenesis QI, increasing the confidence in the tentative identification.

Marker 2 (0.58_175.1188*m/z*, CCS 136.03 Å2) elucidation was carried out in similar way to the process explained, ending in the tentative identification as arginine. The match with the CCS library obtained a delta error of +0.17 %. In this case, an in-source fragment (due to an ammonia neutral loss from the amino group) was also highlighted as marker (0.58_158.0928*m/z*, CCS 132.20 Å2).

As arginine is a compound also susceptible to be ionised under negative ESI mode, the identity confidence was increased by their detection in an additional injection with negative ionization mode ([M-H]$^-$ = 173.1043 *m/z*, retention time 0.59 min, 135.29 Å^2) of one sample from each sampling area following the same chromatographic conditions.

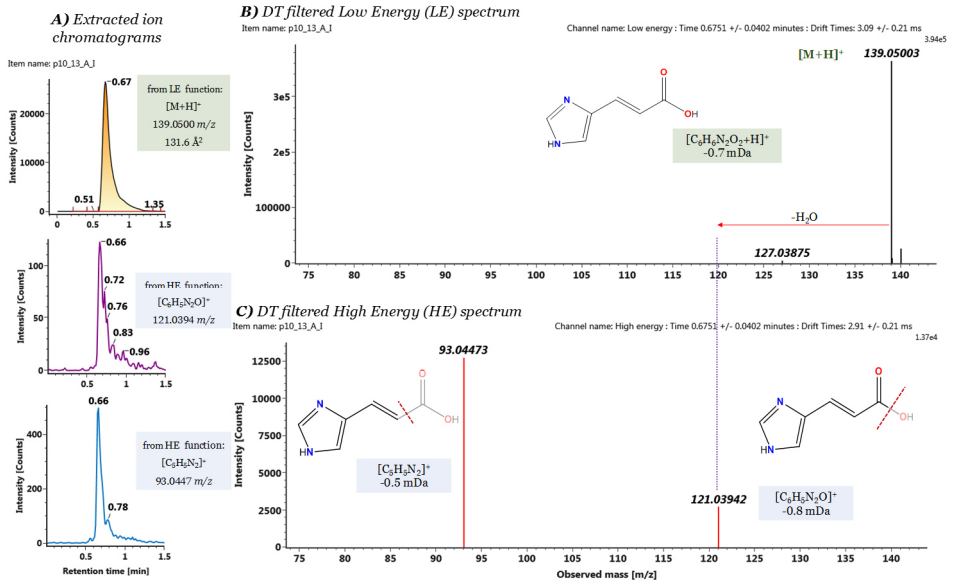


Figure. 7. *Elucidation of Marker 1 Urocanic acid based on the chromatograms and HDMS^E spectra:* ***A)*** *Extracted ion chromatogram from the LE for the parent ion and HE for both fragments along with their experimental m/z and CCS,* ***B)*** *DT filtered LE spectrum,* ***C)*** *DT filtered HE spectrum. A structure fragmentation is proposed.*

The elucidation of markers 3 (11.86_285.2743n, CCS 192.29 Å^2), 4 (12.38_299.2822n, CCS 195.35 Å^2), and 5 (12.67_301.2980n, CCS 197.60 Å^2) was the same as explained above. These three markers were tentatively identified as three free-sphingoid bases, compounds that are formed by a long aliphatic chain with a polar 2-amino-1,3-diol terminus (**Figure S1**). All three showed a similar fragmentation pattern, conformed by 3 fragments due to common neutral loses -H_2O (loss of one hydroxyl group), -2 x H_2O (loss both hydroxyl groups) and the water loss with consecutive formaldehyde (HCHO) loss. Marker 4 was tentatively identified as Sphingosine (d18:1) and marker 5 as Sphinganine (d18:0). For marker 3, three isomeric candidates were possible: erythro-sphingosine (d17:1), (4,14-methy-d16:1) sphingosine and (4,15-methy-d16:1) sphingosine (**Figure S1**). The in-source loss of H_2O ([M-H_2O+H]$^+$) (set as possible ion in the Progenesis PQI deconvolution settings) was present in all three markers, being the predominant ion in the LE spectra in the case of unsaturated markers 3 and 4. For sphinganine (d18:0) (marker 5) a delta CCS error of +0.40 % was obtained with the match

in experimental library. In the other two (3 and 4),), CCS values were predicted using a tool specialised in lipids, LipidCCS [22]. This model was optimised by measuring CCS of a large set of lipids and the estimated error in the prediction for Travelling Wave Ion Mobility (TWIMS) with CCS polyAla calibrants, as VION®, is above ±5 % delta error. This tool allowed to rule out (4,14-methy-d16:1)-sphingosine and (4,15-methy-d16:1)-sphingosine as candidates, since delta CCS error found exceeded -30 %, and select erythro-sphingosine (d17:1) with a delta CCS error of -2.7 % as tentative identification for marker 3. Sphingosine (d18:1) (marker 4) obtained a delta CCS error of -1.1 %.

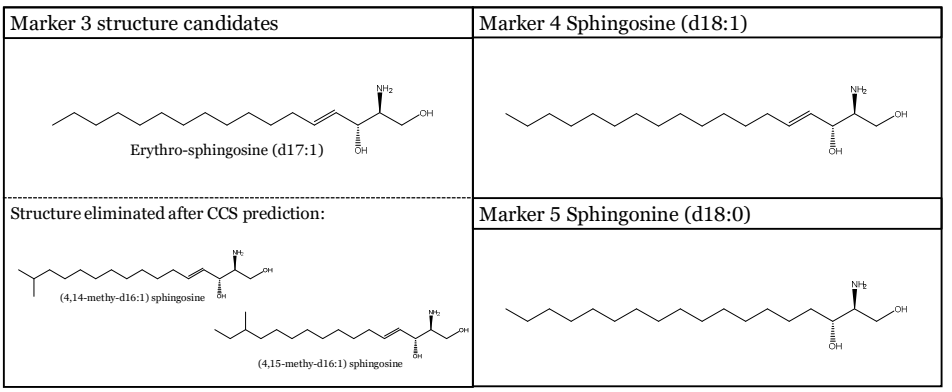


Figure. S1. *Tentative structures for markers 3, 4 and 5, free sphingoid bases.*

Regarding marker 6 (1.88_226.0842n, CCS 145.21 Å2), it is one of the most significative markers for the first two weeks of life due to its high *p(corr)* and high intensity in all the OPLS-DA facings (**Figure 5** and **6**). In addition to the protonated and sodium adducts, sodiated dimer was also detected. Moreover, 4 features highlighted in the statistical analysis were associated to this compound as in-source fragments (1.87_125.0595*m/z*, CCS 123.65 Å^2; 1.87_170.0569*m/z*, CCS 129.58 Å^2; 1.88_153.0544*m/z*, CCS 125.91 Å^2; and 1.88_167.0701*m/z*; CCS 145.21 Å^2). The HDMS^E spectra are shown in **Figure 8**, where some of the highlighted in-source fragments can be observed in the DT filtered LE spectra. The fact that these fragments highlighted as markers, with different CCS and therefore DT from the parent molecule, can also be seen in the filtered LE spectra, shows the easy

fragmentation with little energy of these molecule not only prior to the IMS (giving different CCS) but also post IMS in the collision cell with 6 eV (associated to the CCS of the parent ion). Consequently, all of these fragments are also observed in the filtered HE spectrum, along with many other fragments. In fact, this molecule presents a very rich fragmentation spectrum, which together with the isotopic pattern and the adducts found, renders a definitive molecular formula as $C_{11}H_{14}O_5$ (+0.3 ppm error). Many similar compounds obtained (**Figure S2**) an acceptable match when comparing the DT filtered HE spectrum with the compound/spectral libraries (Metlin, HMDS, Metabolic Profiling CCS Library form Progenesis QI, among others libraries from ChemSpider specialised in metabolites). Above all, a definitive identification could not be given to this compound. On one hand, some fragments were present in the spectra of several candidates (some of them predicted spectra, other experimental) making difficult to decide on one of them. On other hand, other fragments contributed to identification uncertainty as their presence could not be justified. Fragments *m/z* 77.0386 ($[C_6H_5]^+$, -0.5 mDa error), *m/z* 79.0546 ($[C_6H_7]^+$, -0.2 mDa error) and *m/z* 91.0544 ($[C_7H_7]^+$, -0.3 mDa error, tropylium ion) are known to be present in compounds with an bencilic group (C_6H_5-CH_2-). With this, 4 out of 5 unsaturations attributed to the aforementioned molecular formula would be from a benzene ring, leaving an unsaturation probably as a carbonyl group. The presence of the fragment *m/z* 195.0650 (-CH_4O neutral loss, -0.3 mDa error) in DT filtered LE spectrum could be attributed to some methoxy groups with a methyl (some of the candidates with this group presented this fragment in their spectra). The relatively high polarity of this compound (retention time 1.88 min) would suggest the presence of several alcohol groups that could be attached to the aromatic ring or to an attached alkyl chain. However, this would cause the presence of fragments due to water neutral losses not observed in the fragmentation spectra. In the additional analysis of selected samples in negative ESI mode, no peak was observed at the same retention time, therefore it could not be confirmed the presence of a phenol or acid group (negatively ionizable).

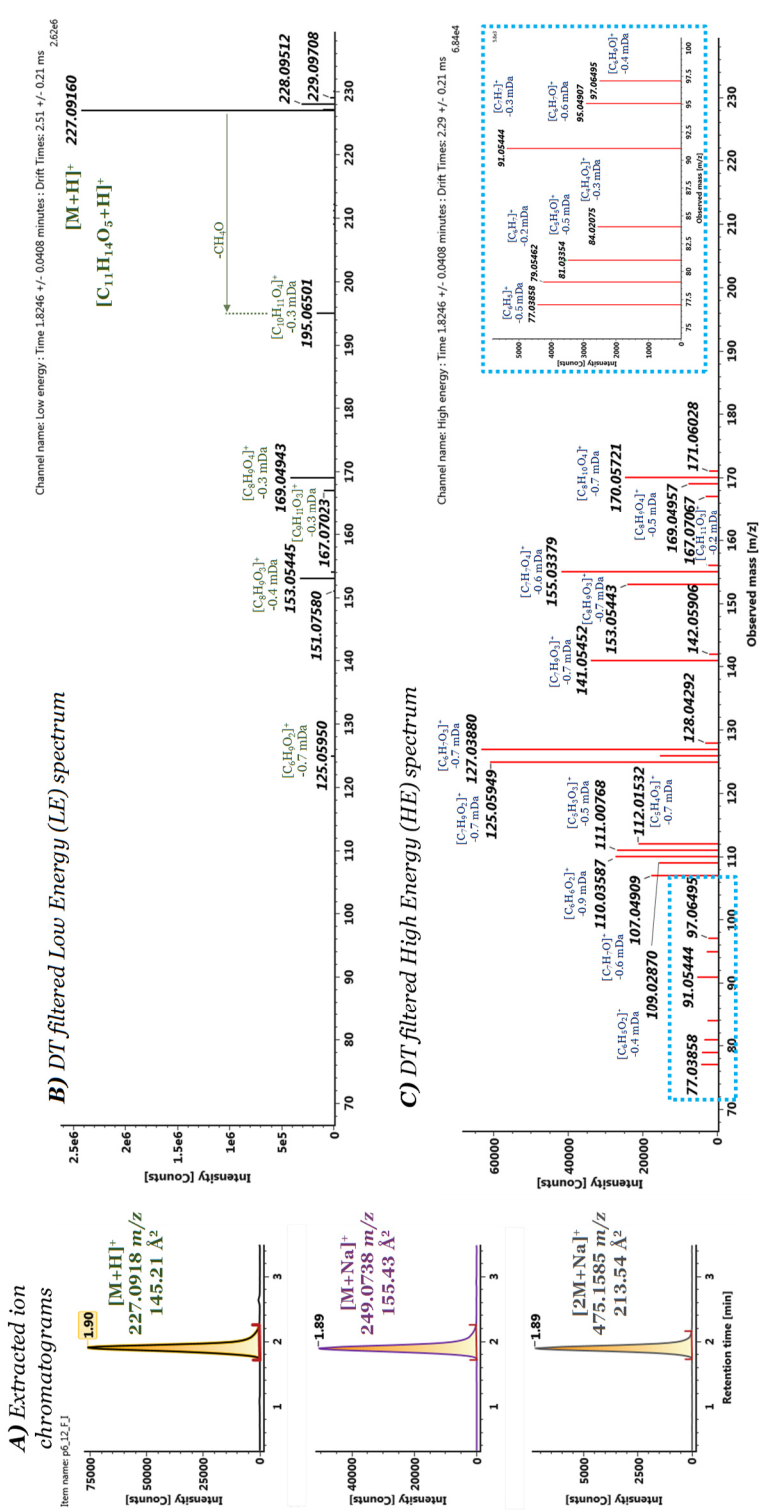


Figure. 8. *Marker 6 the chromatograms and HDMS^E spectra:* ***A)*** *Extracted ion chromatogram from the LE for the adducts of parent compound along with their experimental m/z and CCS,* ***B)*** *DT filtered LE spectrum,* ***C)*** *DT filtered HE spectrum.*

Additionally, as an additional tool to try to elucidate marker 6, it was employed an in-house Rt and CCS prediction tool in our group by Celma et al 2021 [23] (previous work in CCS prediction modelling Bijlsma et al, 2017 [19]) for some of the candidate structures (**Figure S2**). The relative errors obtained with this prediction tool were below ±4.05 % for $[M+H]^+$ and ±5.59 % for $[M+Na]^+$, and ± 2.32 min for the retention with 95 % confidence intervals. It was observed that compounds with many methoxy groups (such as ethyl syringate, 3,4,5-trimethoxyphenyl acetate or 2,4,6-trimethoxyphenyl acetate), showed predicted CCS errors within the accepted limits of the model, but they presented a predicted retention time with a deviation of approximately 4 min, above the accepted error for this parameter. Thus, no tentative identification could be attribute to the candidates selected for the prediction, as none of them get an accepted error for all three parameters predicted. Ultimately, this compound is awaiting further experiments to determine its identity, as HDMS/MS experiments, since due to our group policy for identifications, none of the candidates can be accepted with the current data.

With the 6 markers recorded in **Table 3**, 11 of the 19 highlighted markers have been covered, considering the in-source fragments of markers 1 and 6.

3.4. Putative low-volatile pheromones

As part of the confirmation as a marker of a tentatively identified compound, a bibliographic search is made on its plausibility as a pheromone. This step is especially important to justify the role of the marker in the metabolomic studies carried out. For example, among the possible candidates for marker 2 (finally tentatively identified as urocanic acid) was nicoxamat, an uricosuric drug that in no case would have been given to mice, therefore, it cannot be the identity of the compound, much less a pheromone.

Figure 9 shows the variation in abundance across weeks of the markers (**Table 3**) and the comparison between facial and anogenital sampling area. As it can be observed, all the markers were present in the first two weeks of life in samples from facial and anogenital area. An explanation could be that as part of the care that mice dams provide to their pups, there

is a constant cleaning by licking. This could cause the distribution of the compounds from one area to the other. In addition, this fact could be also explained if these compounds are released in both areas in aqueous secretions, saliva or tears for the facial area and urine for anogenital area. Alternatively, these compounds are products of the metabolism of keratinocytes, which do not differ between regions of the body. For instance, urocanic acid is a common constituent of the stratum corneum of the epidermis of the skin of mammals with possible photoprotective effect [24]. A simple explanation of why it is enriched in the samples from neonatal pups could be that they are hairless, and this makes the skin outer strata easier to erode even with gentle rubbing of the skin with a cotton swab. Moreover, urocanic acid is a metabolite product of histidine metabolism, particularly abundant in mammalian skin secretion and also present in saliva and urine. Its relevance in chemical communication has been already noticed as an attractant in mammalian skin for parasitic nematodes [25].

Arginine is an amino acid that plays and important role in many metabolic reactions and body functions as cell division, immune functions and protein production among others. It can be found in both urinary excretion and saliva secretion. Due to its high presence in the body, it is difficult to know if it has a role as pheromone. Although it is widely known that small peptides and bigger proteins can act as such in mice [26], it is unlikely that a simple amino acid can be used for intraspecies communication

Regarding markers 3, 4 and 5, free-sphingoid bases as sphingosine isomers and sphinganine are employed in different lipids synthesis as ceramides and other sphingolipids that serve as structural and signalling molecules in various cellular events [27]. They are present in all living species ranging from bacteria to humans, and in mammals can be found on the skin (secreted via sebaceous glands), saliva, urine and faeces [28]. Due to their high presence in animals, it is difficult to think that these compounds are pheromones by themselves, since intra-specie communication requires specific molecules that are not easily found in other organisms in the surroundings that could interfere with the communication.

Table 3. *Compound list obtained from the untargeted metabolomic approach: tentative identification and statistical parameters.*

No.	Identification level [a]	Compound	Elemental composition	*p(corr)*	Marker of	Feature name	Rt (min)	protonated molecule *m/z*	Mass error (mDa/ppm)	Experimental CCS (Å²) protonated molecule *m/z*	CCS delta error (%)	Adducts detected
1	2a	Urocanic acid	$C_6H_6N_2O_2$	0.91	W1, Facial W1 vs W4	0.67_139.0500m/z	0.67	139.0500	-0.2/-1.4	131.15	+1.35 %[b]	$[M+H]^+$
				0.90	W2, Facial W2 vs W4							
				0.88	W1, Anogenital W1 vs W4							
				0.85	W2, Anogenital W2 vs W4							
				0.61	W2, Anogenital W1 vs W2							
2	2a	Arginine	$C_6H_{14}N_4O_2$	0.95	W1, Facial W1 vs W4	0.58_175.1188m/z	0.58	175.1188	-0.2/-1.1	136.03	+0.17 %[b]	$[M+H]^+$
				0.92	W2, Facial W2 vs W4							
				0.90	W1, Anogenital W1 vs W4							
				0.89	W2, Anogenital W2 vs W4							
3	3	Erythro-sphingosine – (d17:1)	$C_{17}H_{35}NO_2$	0.88	W1, Facial W1 vs W4	11.86_285.2743n	11.86	286.2743	-0.3/+1.0	192.29	-2.7 %[c]	$[M+H]^+$ $[M-H_2O+H]^+$
				0.89	W1, Anogenital W1 vs W4							
				0.75	W2, Anogenital W2 vs W4							
4	2b	Sphingosine (d18:1)	$C_{18}H_{37}NO_2$	0.90	W1, Facial W1 vs W4	12.38_299.2822n	12.38	300.2901	-0.2/-0.5	195.35	-1.1 %[c]	$[M+H]^+$ $[M-H_2O+H]^+$
				0.70	W2, Facial W2 vs W4							
				0.89	W1, Anogenital W1 vs W4							
				0.79	W2, Anogenital W2 vs W4							
5	2a	Sphinganine (d18:0)	$C_{18}H_{39}NO_2$	0.87	W1, Facial W1 vs W4	12.67_301.2980n	12.67	302.3053	-0.1/-0.3	197.60	+0.40 %[b]	$[M+H]^+$ $[M-H_2O+H]^+$
				0.87	W1, Anogenital W1 vs W4							
				0.76	W2, Anogenital W2 vs W4							
6	4	Unknown	$C_{11}H_{14}O_5$	0.95	W1, Facial W1 vs W4	1.88_226.0842n	1.88	227.0916	-0.4 /+1.5	145.21		$[M+H]^+$ $[M+Na]^+$ $[2M+Na]^+$
				0.91	W2, Facial W2 vs W4							
				0.87	W1, Anogenital W1 vs W4							
				0.86	W2, Anogenital W2 vs W4							

[a] *Identification level according to Celma et al. (2020)*

[b] *CCS delta error (%) by comparison with the experimental CCS from Metabolic Profiling CCS library for Progenesis QI (measured in Waters TWIMS instruments)*

[c] *CCS delta error (%) by comparison with the predicted CCS for protonated molecule obtained by LipidCCS prediction tool developed by Zhou et al. (2017)*

Since an identification for marker 6 has not been reached, it is difficult to determine its plausibility as pheromone. However, compounds with a structure similar to that described in section 3.3 for marker 6 have been reported as pheromones of different organisms (i.e. 3,4-dimethoxycinnamyl alcohol, 4,5-dimethoxyphenol or the isomeric compound 2-hydroxy-3,4,6-trimethoxy-acetophenone found in www.pherobase.com). This demonstrate the potential of marker 6 as a chemical cue and, therefore, its identification will be a target in future research, either by means of the injection of different analytical standards of different candidates or by the use of other analytical techniques.

Therefore, with the exception of marker 6 that awaits identification, the remaining markers seem very common metabolites present in the skin of mammals and/or associated to secretion of exocrine glands (salivary, lacrimal, sebaceous). Although this make them unlikely candidate to pheromones by themselves, when present together in a certain ratio, they may constitute a pheromonal blend [2], a multicomponent pheromone characterising the hairless skin of neonatal pups and facilitating maternal care (e.g. licking grooming). This deserves being tested in the near future.

4. Conclusions

Following the first part of the study focused on volatile pheromones, in this work we have successfully applied an untargeted metabolomic approach to discovery of mouse pup low-volatile part of pheromones. Sample collection with swabs allowed the work with alive animals at different stages of their development without interfering in their normal growths and welfare. The identification of putative low-volatile pheromones involved in the maternal care in mice was possible thanks to the combination of untargeted metabolomics with UHPLC-IMS-HRMS. As in previous work, the differentiating compounds when facing samples from the fourth week of mouse life (time of weaning) against the sample from neonatal mice (first two weeks, receiving maternal care), essentially belonged to the latter. This supports previous studies that suggest the role of chemical communication in maternal care and the hypothesis that this behaviour is induced by compound emitted in the first weeks of life. Five candidates to pup pheromones were tentatively identified, and a sixth compound needs further investigation to

give a candidate structure. All of them had a great presence in neonatal pups and shows lower abundance in 4-week youngsters. The introduction of IMS to the well stablished LC-HRMS instruments provided additional structural information that, in combination with CCS, compound/spectral databases and CCS predictors models; facilitates the compound identification of most of the compiled markers. Further behavioural research with the tentatively elucidated compounds will be carried out in order to clarify their role in chemical communication for the maternal behaviour induction.

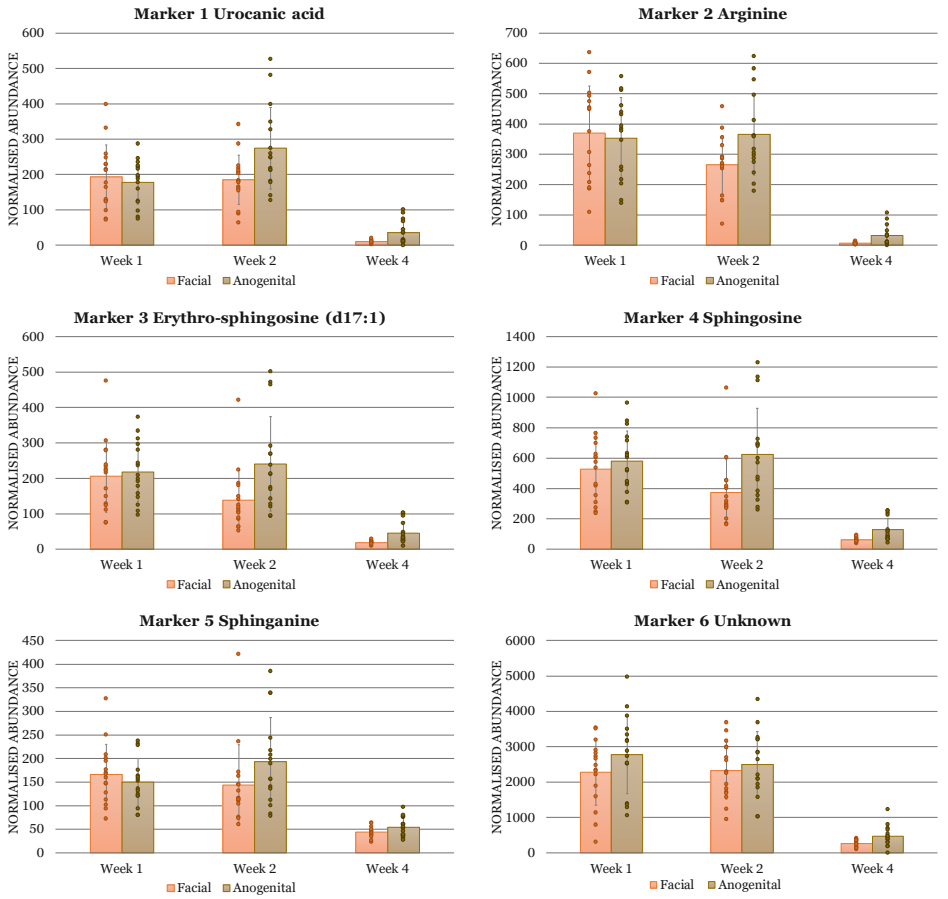


Figure. 8. *Comparison of normalised abundance between facial and anogenital areas sampling areas, for the 6 markers. In orange shades the average of normalised abundance obtained from the samples of the same week from facial sampling area (■) and brown shades from anogenital sampling area (■).*

Deontological

These experiments were performed throughout following the European Union Council Directive of June 3rd, 2010 (6106/1/10 REV1). Accordingly, procedures were approved by the Committee of Ethics on Animal Experimentation of the Jaume I University of Castellón where the experiments were performed and, ultimately, by the Valencian *Conselleria d'Agricultura Medi Ambient, Canvi Climàtic i Desenvolupament Rural* (code 2019/VSC/PEA/0049).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

L Lacalle-Bergeron acknowledges the financial support of Universitat Jaume I, Spain for his pre-doctoral grant (UJI 19I001/03). T. Portoles acknowledges Ramon y Cajal Program from the Ministry of Economy and Competitiveness, Spain (RYC-2017-22525) for funding her research. The Research Institute for Pesticides and Water (IUPA) authors acknowledge the financial support of Generalitat Valenciana, as research group of excellence PROMETEO/2019/040 and Universitat Jaume I de Castelló (UJI-B2020-25 and UJI-B2020-37). The Laboratory of Functional Neuroanatomy (Unitat Mixta NeuroFun-UV-UJI) authors acknowledge for financial support of Generalitat Valenciana, as research group of excellence PROMETEO/2017/078, the Spanish Ministry of Science and Innovation (BFU2016-77691-C2; PID2019-107322GB-C21) and Universitat Jaume I de Castelló (UJI-B2016-45).

References

[1] P. Karlson, M. Lüscher, 'Pheromones': a new term for a class of biologically active substances, Nature. 183 (1959) 55–56. https://doi.org/10.1038/183055a0.

[2] T.D. Wyatt, Pheromones and signature mixtures: defining species-wide signals and variable cues for identity in both invertebrates and

vertebrates, J. Comp. Physiol. A. 196 (2010) 685–700. https://doi.org/10.1007/s00359-010-0564-y.

[3] M. Meredith, R.J. O'Connell, Efferent control of stimulus access to the hamster vomeronasal organ., J. Physiol. 286 (1979) 301–316. https://doi.org/10.1113/jphysiol.1979.sp012620.

[4] Y. Isogai, S. Si, L. Pont-Lezica, T. Tan, V. Kapoor, V.N. Murthy, C. Dulac, Molecular organization of vomeronasal chemoreception, Nature. 478 (2011) 241–245. https://doi.org/10.1038/nature10437.

[5] T.D. Wyatt, Semiochemicals and social organization, in: Pheromones Anim. Behav., Cambridge University Press, Cambridge, 2014: pp. 126–149. https://doi.org/10.1017/CBO9781139030748.008.

[6] Y. Isogai, Z. Wu, M.I. Love, M.H.-Y. Ahn, D. Bambah-Mukku, V. Hua, K. Farrell, C. Dulac, Multisensory logic of infant-directed aggression by males, Cell. 175 (2018) 1827-1841.e17. https://doi.org/10.1016/j.cell.2018.11.032.

[7] T.S. Nakahara, L.M. Cardozo, X. Ibarra-Soria, A.D. Bard, V.M.A. Carvalho, G.Z. Trintinalia, D.W. Logan, F. Papes, Detection of pup odors by non-canonical adult vomeronasal neurons expressing an odorant receptor gene is influenced by sex and parenting status, BMC Biol. 14 (2016) 12. https://doi.org/10.1186/s12915-016-0234-9.

[8] J.G. Vandenbergh, Effects of central and peripheral anosmia on reproduction of female mice, Physiol. Behav. 10 (1973) 257–261. https://doi.org/10.1016/0031-9384(73)90307-7.

[9] Z. Wang, D.R. Storm, Maternal behavior is impaired in female mice lacking type 3 adenylyl cyclase, Neuropsychopharmacology. 36 (2011) 772–781. https://doi.org/10.1038/npp.2010.211.

[10] R.S. Bridges, Neurobiology of the parental brain, Elsevier, 2008. https://doi.org/10.1016/B978-0-12-374285-8.X0001-7.

[11] L. Lacalle-Bergeron, R. Goterris-Cerisuelo, T. Portolés, J. Beltran, J.V. Sancho, C. Navarro-Moreno, F. Martinez-Garcia, Novel sampling strategy for alive animal volatolome extraction combined with GC-MS based untargeted metabolomics: Identifying mouse pup pheromones,

Talanta. 235 (2021) 122786. https://doi.org/10.1016/j.talanta.2021.122786.

[12] K. Dettmer, P.A. Aronov, B.D. Hammock, Mass spectrometry-based metabolomics, Mass Spectrom. Rev. 26 (2007) 51–78. https://doi.org/10.1002/mas.20108.

[13] L. Lacalle-Bergeron, D. Izquierdo-Sandoval, J.V. Sancho, F.J. López, F. Hernández, T. Portolés, Chromatography hyphenated to high resolution mass spectrometry in untargeted metabolomics for investigation of food (bio)markers, TrAC Trends Anal. Chem. 135 (2021) 116161. https://doi.org/10.1016/j.trac.2020.116161

[14] M. Bevilacqua, R. Bro, F. Marini, Å. Rinnan, M.A. Rasmussen, T. Skov, Recent chemometrics advances for foodomics, TrAC Trends Anal. Chem. 96 (2017) 42–51. https://doi.org/10.1016/j.trac.2017.08.011.

[15] X. Zhang, Q. Li, Z. Xu, J. Dou, Mass spectrometry-based metabolomics in health and medical science: a systematic review, RSC Adv. 10 (2020) 3092–3104. https://doi.org/10.1039/C9RA08985C.

[16] M.-D. Luo, Z.-W. Zhou, Z.-J. Zhu, The Application of Ion Mobility-Mass Spectrometry in Untargeted Metabolomics: from Separation to Identification, J. Anal. Test. 4 (2020) 163–174. https://doi.org/10.1007/s41664-020-00133-0.

[17] C.M. Nichols, J.N. Dodds, B.S. Rose, J.A. Picache, C.B. Morris, S.G. Codreanu, J.C. May, S.D. Sherrod, J.A. McLean, Untargeted Molecular Discovery in Primary Metabolism: Collision Cross Section as a Molecular Descriptor in Ion Mobility-Mass Spectrometry, Anal. Chem. 90 (2018) 14484–14492. https://doi.org/10.1021/acs.analchem.8b04322.

[18] C.B. Mollerup, M. Mardal, P.W. Dalsgaard, K. Linnet, L.P. Barron, Prediction of collision cross section and retention time for broad scope screening in gradient reversed-phase liquid chromatography-ion mobility-high resolution accurate mass spectrometry, J. Chromatogr. A. 1542 (2018) 82–88. https://doi.org/10.1016/j.chroma.2018.02.025.

[19] L. Bijlsma, R. Bade, A. Celma, L. Mullin, G. Cleland, S. Stead, F. Hernandez, J. V. Sancho, Prediction of Collision Cross-Section Values for

Small Molecules: Application to Pesticide Residue Analysis, Anal. Chem. 89 (2017) 6583–6589. https://doi.org/10.1021/acs.analchem.7b00741.

[20] B. Worley, R. Powers, Multivariate Analysis in Metabolomics, Curr. Metabolomics. 1 (2012) 92–107. https://doi.org/10.2174/2213235X130108.

[21] A. Celma, J. V. Sancho, E.L. Schymanski, D. Fabregat-Safont, M. Ibáñez, J. Goshawk, G. Barknowitz, F. Hernández, L. Bijlsma, Improving Target and Suspect Screening High-Resolution Mass Spectrometry Workflows in Environmental Analysis by Ion Mobility Separation, Environ. Sci. Technol. (2020). https://doi.org/10.1021/acs.est.0c05713.

[22] Z. Zhou, J. Tu, X. Xiong, X. Shen, Z.-J. Zhu, LipidCCS: prediction of collision cross-section values for lipids with high precision to support ion mobility–mass spectrometry-based lipidomics, Anal. Chem. 89 (2017) 9559–9566. https://doi.org/10.1021/acs.analchem.7b02625.

[23] A. Celma, R. Bade, J.V. Sancho, F. Hernández, M. Humphries, L. Bijlsma, Dual prediction of retention time and collision cross section of emergin contaminants for environmental analyses, Analytical Chemistry (2021) Submitted

[24] N.K. Gibbs, M. Norval, Urocanic acid in the skin: a mixed blessing?, J. Invest. Dermatol. 131 (2011) 14–17. https://doi.org/10.1038/jid.2010.276.

[25] D. Safer, M. Brenes, S. Dunipace, G. Schad, Urocanic acid is a major chemoattractant for the skin-penetrating parasitic nematode *Strongyloides stercoralis*, Proc. Natl. Acad. Sci. 104 (2007) 1627–1630. https://doi.org/10.1073/pnas.0610193104.

[26] M.M. Phelan, L. McLean, S.D. Armstrong, J.L. Hurst, R.J. Beynon, L.-Y. Lian, The structure, stability and pheromone binding of the male mouse protein sex pheromone darcin, PLoS One. 9 (2014) e108415. https://doi.org/10.1371/journal.pone.0108415.

[27] M.A. Al Sazzad, T. Yasuda, M. Murata, J.P. Slotte, The long-chain sphingoid base of ceramides determines their propensity for lateral

segregation, Biophys. J. 112 (2017) 976–983. https://doi.org/10.1016/j.bpj.2017.01.016.

[28] C.L. Fischer, Antimicrobial activity of host-derived lipids, Antibiotics. 9 (2020) 75. https://doi.org/10.3390/antibiotics9020075.

Supplementary information

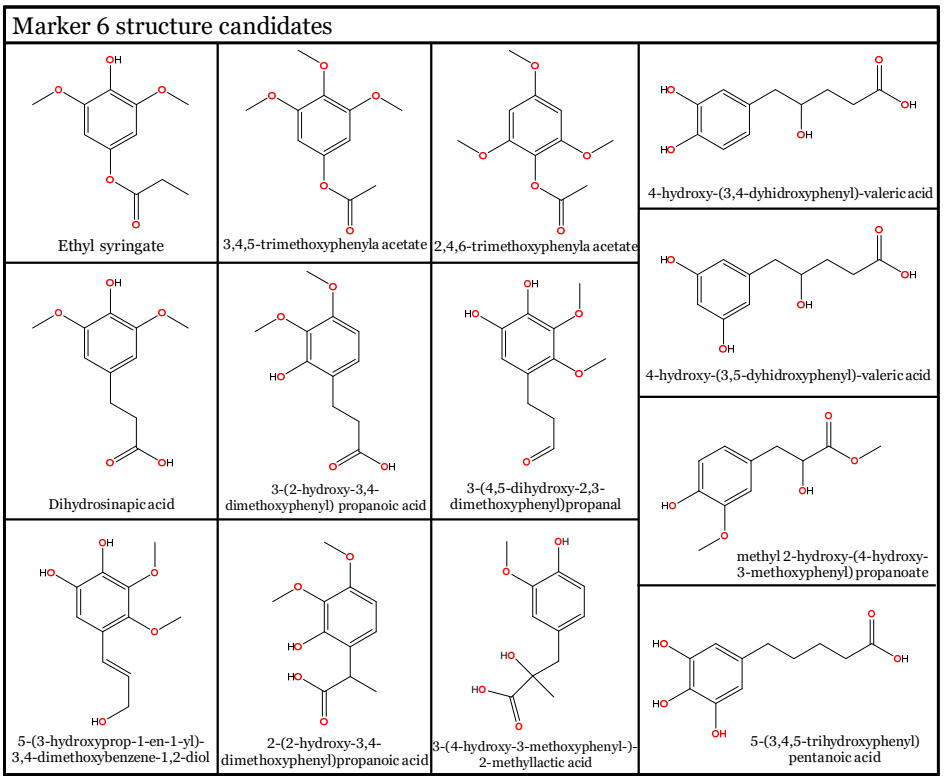


Figure. S2. *Candidates structures for marker 6.*

5.4 Discussion of the results

In this chapter, two different analytical methodologies were employed with untargeted metabolomics for the identification of putative chemosignals related to maternal care in mice. For both studies, the starting hypothesis is that during the first days of age some compounds released by pups could induce in an intensive care by mouse dams as chemical communication response. As the pups grow, the presence of these chemical cues decreases until their weaning.

In both studies, the sampling had to be adapted to the work with alive animals, without affecting their well-being and growth. This was especially important since it was intended to sample the same animals repetitively during the period of time between 4 and 24 days-old, in order to see changes in the secreted/excreted metabolome. It was also important that the process was the less stressful as possible for the pups, since this could lead to the presence of certain compounds associated with stress that could distort the results.

In **Scientific article V**, the focus was on those compounds on a volatile nature that are emitted by pups. For its collection, an adaptation of the DHS-P&T was employed, a technique previously used in our laboratory for the analysis of volatiles in food products, as seen in **Chapter 2** of this thesis. The main change in technique would be the replacement of N_2 flow that is used to carry the compounds to the sorbent. Instead, an air flux was induced by a vacuum pup, allowing the pups to breathe normally. On the other hand, the flasks containing the pups were adapted to their size and, for the smallest ones, a soft thermal sand bath was placed to maintain a comfortable temperature. This methodology allowed to capture the volatolome of whole alive animals, being able to be used in the future with other small animals and adapted for larger animals.

The second study, **Scientific Article VI**, aimed on the obtention of less volatile compounds not covered in the first study. Consecutively to the volatolome extraction, surface areas of the pups body susceptible to be zones with a high release of compound acting as chemosignals (due to the presence

of glands and excretory ducts) were gently rubbed with swabs for their collection.

Regarding the analysis of the compounds, the volatile compounds were analysed by GC-EI-MS using a single quadrupole mass analyser, while for the compounds with lower volatility, LC coupled to IMS-HRMS was employed.

On the one hand, the chromatographic separation technique was adapted to the nature of the compounds of interest in each study, being GC the gold standard for volatile, non-polar and thermostable compounds and LC for polar, low-volatility and thermolabile compounds. On the other hand, mass analysers with very different characteristics have been employed.

GC-MS with EI is one of the oldest and well-stablished coupling techniques, being commercially available for many decades. EI is the ionization mode most employed due to its universality, reproducibility and robustness. Considered a hard-ionisation technique due the high fragmentation of the compounds employing a standardised electron energy of 70 eV, the very reproducibly fragmentation patterns are used for the automated identification of analytes by means of spectrum libraries. The availability of well-stocked spectral libraries as NIST library and the well-stablished methods for retention index calculation (Kovats index and variants as linear retention index) provides a clear advantage for GC-EI-MS in the biomarkers identification in untargeted metabolomics. However, the identifications are based practically on matches with the spectral library, and in case of having doubts about the candidate, excessive fragmentation, and even the absence of the molecular ion, make definitive identification difficult. Furthermore, in this case, since quadrupole mass analyser with nominal mass was employed, the interpretation of certain spectra was difficult.

Conversely, for the analysis of low-volatile putative pheromones, LC is coupled to latest generation of HRMS instrumentation with IMS. In these instruments the peak capacity and dynamic range have been increased, as well as the improvement of the performance over classical MS^E acquisition method and the access to additional structural information about the detected metabolites through the CCS value, as it has been extensively explained in different parts of this thesis.

Despite the fact that IMS represents a great additional help in the identification of compounds, the process of compound elucidation continues to be the main handicap in untargeted metabolomics studies, as it has been shown in **Scientific article VI** with the uncompleted identification of marker 6. However, we are confident that this bottleneck could be widened in the future by increasing the experimental CCS and spectral databases, robust prediction models and the continuous improvement of analytical instruments.

All in all, in this chapter complementary analytical techniques have been employed to cover a large number of compounds that may have a role in chemical communication between mouse pup and dams. The result was the identification of 11 volatile compounds and 5 tentative identifications of less-volatile markers, most of them exalted in the younger mouse pups. The next step is in charge of the research group of Laboratory of Functional Neuroanatomy from *Universitat Jaume I*, whose collaboration has been key for the development of the present works. Further behavioural studies will be performed to clarify the role of putative pheromonones/chemosignals in chemical communication related to maternal behaviour induction of above-mentioned markers.

5.5 References

Brennan, P. (2009). Pheromones and Mammalian Behavior. In The Neurobiology of Olfaction (Issue 1959, pp. 157–179). https://doi.org/10.1201/9781420071993-c6

Bridges, R. S. (2008). Neurobiology of the parental brain. In Neurobiology of the Parental Brain. Elsevier. https://doi.org/10.1016/B978-0-12-374285-8.X0001-7

Butcher, R. A. (2017). Small-molecule pheromones and hormones controlling nematode development. Nature Chemical Biology, 13(6), 577–586. https://doi.org/10.1038/nchembio.2356

Esteves, F. G., Santos-Pinto, J. R. A. dos, Saidemberg, D. M., & Palma, M. S. (2017). Using a proteometabolomic approach to investigate the role of Dufour's gland in pheromone biosynthesis in the social wasp Polybia paulista. Journal of Proteomics, 151, 122–130. https://doi.org/10.1016/j.jprot.2016.01.009

Falcke, J. M., Bose, N., Artyukhin, A. B., Rödelsperger, C., Markov, G. V., Yim, J. J., Grimm, D., Claassen, M. H., Panda, O., Baccile, J. A., Zhang, Y. K., Le, H. H., Jolic, D., Schroeder, F. C., & Sommer, R. J. (2018). Linking Genomic and Metabolomic Natural Variation Uncovers Nematode Pheromone Biosynthesis. Cell Chemical Biology, 25(6), 787-796.e12. https://doi.org/10.1016/j.chembiol.2018.04.004

Fiorini, F., Borgonuovo, C., Ferrante, M. I., & Brönstrup, M. (2020). A metabolomics exploration of the sexual phase in the marine diatom pseudo-nitzschia multistriata. Marine Drugs, 18(6). https://doi.org/10.3390/md18060313

Fortes-Marco, L., Lanuza, E., & Martinez-Garcia, F. (2013). Of Pheromones and Kairomones: What receptors mediate innate emotional responses? The Anatomical Record, 296(9), 1346–1363. https://doi.org/10.1002/ar.22745

Fujita, A., Okuno, T., Oda, M., & Kato, K. (2020). Urinary volatilome analysis in a mouse model of anxiety and depression. PLOS ONE, 15(2), e0229269. https://doi.org/10.1371/journal.pone.0229269

Hasen, N. S., & Gammie, S. C. (2011). Trpc2-deficient lactating mice exhibit altered brain and behavioral responses to bedding stimuli. Behavioural Brain Research, 217(2), 347–353. https://doi.org/10.1016/j.bbr.2010.11.002

Hefetz, A. (2019). The critical role of primer pheromones in maintaining insect sociality. Zeitschrift Für Naturforschung C, 74(9–10), 221–231. https://doi.org/10.1515/znc-2018-0224

Hosseini, S. A., Goldansaz, S. H., Groot, A. T., Menken, S. B. J., Van Der Wielen, F., Wissel, C., Vercammen, J., De Rijke, E., & Roessingh, P. (2021). Identification of bioactive plant volatiles for the carob moth by means of GC-EAD and GC-Orbitrap MS. Applied Sciences, 11(18), 8603. https://doi.org/10.3390/app11188603

Jones, M. K., Huff, T. B., Freeman, E. W., & Songsasen, N. (2021). Differential expression of urinary volatile organic compounds by sex, male reproductive status, and pairing status in the maned wolf (Chrysocyon brachyurus). PLoS ONE, 16(8 August), 1–21. https://doi.org/10.1371/journal.pone.0256388

Karlson, P., & Lüscher, M. (1959). 'Pheromones': a new term for a class of biologically active substances. Nature, 183(4653), 55–56. https://doi.org/10.1038/183055a0

Kimchi, T., Xu, J., & Dulac, C. (2007). A functional circuit underlying male sexual behaviour in the female mouse brain. Nature, 448(7157), 1009–1014. https://doi.org/10.1038/nature06089

Kvasnička, A., Friedecký, D., Tichá, A., Hyšpler, R., Janečková, H., Brumarová, R., Najdekr, L., & Zadák, Z. (2021). SLIDE—Novel Approach to Apocrine Sweat Sampling for Lipid Profiling in Healthy Individuals. International Journal of Molecular Sciences, 22(15), 8054. https://doi.org/10.3390/ijms22158054

Le, H. H., Wrobel, C. J. J., Cohen, S. M., Yu, J., Park, H., Helf, M. J., Curtis, B. J., Kruempel, J. C., Rodrigues, P. R., Hu, P. J., Sternberg, P. W., & Schroeder, F. C. (2020). Modular metabolite assembly in Caenorhabditis elegans depends on carboxylesterases and formation of lysosome-related organelles. ELife, 9(1), 1–42. https://doi.org/10.7554/eLife.61886

Li, K., Buchinger, T. J., & Li, W. (2018). Discovery and characterization of natural products that act as pheromones in fish. Natural Product Reports, 35(6), 501–513. https://doi.org/10.1039/C8NP00003D

Ljunggren, J., Borrero-Echeverry, F., Chakraborty, A., Lindblom, T. U. T., Hedenström, E., Karlsson, M., Witzgall, P., & Bengtsson, M. (2019). Yeast volatomes differentially affect larval feeding in an insect herbivore. Applied and Environmental Microbiology, 85(21). https://doi.org/10.1128/AEM.01761-19

Meredith, M., & O'Connell, R. J. (1979). Efferent control of stimulus access to the hamster vomeronasal organ. The Journal of Physiology, 286(1), 301–316. https://doi.org/10.1113/jphysiol.1979.sp012620

Navarro-Moreno, C., Sanchez-Catalan, M. J., Barneo-Muñoz, M., Goterris-Cerisuelo, R., Belles, M., Lanuza, E., Agustin-Pavon, C., & Martinez-Garcia, F. (2020). Pregnancy Changes the Response of the Vomeronasal and Olfactory Systems to Pups in Mice. Frontiers in Cellular Neuroscience, 14(December), 1–19. https://doi.org/10.3389/fncel.2020.593309

Nordlund, D. A., & Lewis, W. J. (1976). Terminology of chemical releasing stimuli in intraspecific and interspecific interactions. Journal of Chemical Ecology, 2(2), 211–220. https://doi.org/10.1007/BF00987744

Numan, M., & Insel, T. R. (2003). The neurobiology of parental behavior (1st ed.). Springer-Verlag New York. https://doi.org/10.1007/b97533

Park, Joo, Park, & Paik. (2019). Ascaroside Pheromones: Chemical Biology and Pleiotropic Neuronal Functions. International Journal of Molecular Sciences, 20(16), 3898. https://doi.org/10.3390/ijms20163898

Smeets, M. A. M., Rosing, E. A. E., Jacobs, D. M., van Velzen, E., Koek, J. H., Blonk, C., Gortemaker, I., Eidhof, M. B., Markovitch, B., de Groot, J., &

Semin, G. R. (2020). Chemical fingerprints of emotional body odor. Metabolites, 10(3), 84. https://doi.org/10.3390/metabo10030084

Vandenbergh, J. G. (1973). Effects of central and peripheral anosmia on reproduction of female mice. Physiology and Behavior, 10(2), 257–261. https://doi.org/10.1016/0031-9384(73)90307-7

Varshavi, D., Scott, F. H., Varshavi, D., Veeravalli, S., Phillips, I. R., Veselkov, K., Strittmatter, N., Takats, Z., Shephard, E. A., & Everett, J. R. (2018). Metabolic biomarkers of ageing in C57BL/6J wild-type and flavin-containing monooxygenase 5 (FMO5)-knockout mice. Frontiers in Molecular Biosciences, 5(APR), 1–16. https://doi.org/10.3389/fmolb.2018.00028

Wyatt, T. D. (2010). Pheromones and signature mixtures: defining species-wide signals and variable cues for identity in both invertebrates and vertebrates. Journal of Comparative Physiology A, 196(10), 685–700. https://doi.org/10.1007/s00359-010-0564-y

Zheng, C., Zhou, Q., Wang, Z., & Wang, J. (2021). Behavioral responses of Platycladus orientalis plant volatiles to Phloeosinus aubei by GC-MS and HS-GC-IMS for discrimination of different invasive severity. Analytical and Bioanalytical Chemistry, 413(23), 5789–5798. https://doi.org/10.1007/s00216-021-03556-5

Chapter 6.
Conclusions and future work

6.1 Conclusions

During the present doctoral thesis, untargeted metabolomic approach has been applied and evaluated with different advanced analytical techniques into various fields of knowledge. On the one hand, three studies focused on different areas related to food area were carried out, ranging from food processing, food intake to the effects of nutrition on health. On the other hand, two different analytical methodologies were employed along with untargeted metabolomics for the discovery of compounds related to intra-species communication, a field of study little addressed with this approach until now. The **general conclusion** of this thesis is that untargeted metabolomics together with the coupling of different chromatographic separation techniques with mass spectrometry analysers are powerful tools for the detection and identification of (bio) markers associated with issues in different fields of study.

The **specific conclusions** of the different studies carried out are detailed below:

- ◇ The technological advances in both chromatography and mass spectrometry instrumentation have been crucial for the impulse of untargeted metabolomics in many research fields in the last years.
- ◇ The efficient separation combined with the acquisition of sensitive and high-quality structural compound information from the hyphenation of modern chromatography and mass spectrometry analysers facilitates the detection and identification of metabolites in complex biological samples.
- ◇ The correct study design and appropriated sample treatments are critical in the development of untargeted metabolomic strategies.
- ◇ The increasing demand for more exhaustive control over food processing, in terms of authenticity, quality and safety, can be addressed by untargeted metabolomics.
- ◇ The analysis of the volatile composition of smoked fish samples extracted by DHS-P&T and the subsequent analysis by GC-MS allowed the development of a classification model to distinguish

samples with "Cold Smoked" treatment from other banned processes in the European Union.

- ◊ The application of untargeted metabolomics using the novel UHPLC-IMS-HRMS has revealed powerful for the identification biomarkers related to the intake of food products and diets, in particular to the discovery of short-term biomarkers of orange intake.
- ◊ The additional separation obtained with IMS provides an additional structural descriptor and cleaner MS spectra highly valuable for the identification process of unknown compounds in untargeted metabolomics.
- ◊ Untargeted metabolomics can help to understand the complex relationships between nutritional exposure and health status, in addition to the study of potentially beneficial food products or supplementation.
- ◊ The potential of UHPLC-IMS-HRMS analysis along with untargeted metabolomics proved with the elucidation of biomarkers related to the treatment with resveratrol and pterostilbene to rat with liver steatosis. The putative benefits for health of these supplementation has been demonstrated based on the compounds tentatively elucidated.
- ◊ The CCS experimental database and CCS prediction tools have been created around the CCS value that have facilitated the compound identification.
- ◊ The untargeted metabolomic approach has proven its usefulness in the determination of putative chemosignals/pheromones for the chemical communication intra-species, specifically for the materno-filial communication in mice.
- ◊ The novel sampling procedure based on DHS-P&T for volatolome extraction of whole, alive mouse pups and the subsequent analysis by GC-MS allowed to identify putative volatile pheromones involved in maternal behaviour.
- ◊ The detection of low-volatile putative mouse pup chemosignals was possible thanks to an UHPLC-IMS-HRMS based untargeted metabolomic approximation.

Last but not least, the importance of multidisciplinary collaboration has been demonstrated throughout this thesis. The application of advanced analytical methodologies to give answer to questions exposed by food industries, experts in nutrition and neurobiologists has allowed to increase the knowledge on these fields.

6.1 Conclusiones

Durante la presente tesis doctoral, se ha aplicado y evaluado la aproximación de la metabolómica no dirigida con distintas técnicas analíticas avanzadas en diferentes campos del conocimiento. Por un lado, se llevaron a cabo tres estudios enfocados a diferentes áreas relacionadas con el campo de la alimentación, que van desde el procesamiento de alimentos, la ingesta de alimentos y hasta los efectos de la nutrición en la salud. Por otro lado, se emplearon dos metodologías analíticas diferentes junto con una metabolómica no dirigida para el descubrimiento de compuestos relacionados con la comunicación intra-especie, un campo de estudio hasta ahora poco abordado con esta metodología. La **conclusión general** de esta tesis es que la metabolómica no dirigida junto con el acoplamiento de distintas técnicas de separación cromatográfica con analizadores de espectrometría de masas son herramientas poderosas a la hora de detectar e identificar (bio)marcadores asociados a problemáticas de distintos ámbitos de estudio.

A continuación, se detallan las **conclusiones específicas** relacionadas con los trabajos realizados:

- ◇ Los avances tecnológicos en los últimos años en la instrumentación de cromatografía y espectrometría de masas han sido cruciales para el impulso de la metabolómica no dirigida en muchos campos de investigación.
- ◇ La separación eficiente combinada con la adquisición de información estructural de alta calidad a partir del acoplamiento de técnicas cromatográficas modernas con potentes analizadores de espectrometría de masas, facilita la detección e identificación de metabolitos en muestras biológicas complejas.
- ◇ El correcto diseño del estudio y la selección de los tratamientos de muestra apropiados son fundamentales para el desarrollo de estrategias metabolómicas no dirigidas.
- ◇ La creciente demanda de un control más exhaustivo del procesamiento de alimentos, en términos de autenticidad, calidad y seguridad, puede abordarse mediante la metabolómica no dirigida.

- ◊ El análisis de los compuestos volátiles de muestras de pescado ahumado extraídas por DHS-P & T y el posterior análisis por GC-MS permitió desarrollar un modelo de clasificación para distinguir muestras con tratamiento "*Cold smoke*" de otros procesos alimentarios prohibidos en la Unión Europea.
- ◊ La aplicación de metabolómica no dirigida utilizando la novedosa técnica UHPLC-IMS-HRMS ha demostrado su gran potencial en la identificación de biomarcadores relacionados con la ingesta de productos alimenticios y dietas específicas, en particular con el descubrimiento de biomarcadores a corto plazo de ingesta de naranja.
- ◊ La metabolómica no dirigida puede ayudar a comprender las complejas relaciones entre la exposición nutricional y el estado de salud, además del estudio de los productos alimenticios o suplementos potencialmente beneficiosos.
- ◊ El potencial del análisis UHPLC-IMS-HRMS junto con la metabolómica no dirigida también se ha confirmado con la elucidación de biomarcadores relacionados con el tratamiento con resveratrol y pterostilbeno en ratas con esteatosis hepática. Los compuestos elucidados tentativamente han demostrado el posible beneficio para la salud que tendría esta suplementación.
- ◊ Las herramientas creadas en torno al valor de CCS como bases de datos experimentales CCS y herramientas de predicción han facilitado la etapa de identificación de compuestos (bio)marcadores.
- ◊ El enfoque de la metabolómica no dirigida ha demostrado su utilidad en la determinación de señales químicas/feromonas putativas en la comunicación química intra-especie, específicamente para la comunicación maternofilial en ratones.
- ◊ El novedoso procedimiento de muestreo basado en DHS-P&T para la captación del volatoloma de crías de ratón vivas y el análisis posterior por GC-MS permitió identificar potenciales feromonas volátiles involucradas en el comportamiento materno.

- La detección de putativas señales químicas de baja volatilidad en crías de ratón fue posible gracias a una aproximación metabolómica no dirigida basada en UHPLC-IMS-HRMS.

Por último, pero no menos importante, a lo largo de esta tesis se ha dejado patente la importancia de la colaboración multidisciplinar. La aplicación de metodologías analíticas avanzadas para dar respuesta a problemáticas expuestas por industrias alimentarias, expertos en nutrición y neurobiólogos han permitido incrementar el conocimiento en estos campos.

6.1 Conclusion

Au cours de cette thèse doctorale, l'approche de la métabolomique non ciblée a été appliquée et évaluée avec différentes techniques analytiques avancées dans différents domaines de la connaissance. D'une part, trois études portant sur différents domaines liés à l'alimentation ont été menées, qui comprend le traitement des aliments, la consommation alimentaire et même les effets de la nutrition sur la santé. D'autre part, deux méthodologies analytiques différentes ont été utilisées avec une métabolomique non ciblée pour la découverte de composés liés à la communication intra-espèce, un domaine d'étude jusqu'à présent peu abordé avec cette méthodologie. La **conclusion générale** de cette thèse est que la métabolomique non ciblée ainsi que le couplage de différentes techniques de séparation chromatographique avec des analyseurs de spectrométrie de masse sont des outils puissants lorsqu'il s'agit de détecter et d'identifier des (bio)marqueurs pour répondre aux questions posées par différents domaines d'étude.

Les **conclusions spécifiques** liées aux travaux réalisés sont détaillées ci-dessous :

- ◇ Les progrès technologiques de ces dernières années dans les instruments de chromatographie et de spectrométrie de masse ont été cruciales pour l'avancement de la métabolomique non ciblée dans de nombreux domaines de recherche.
- ◇ Une séparation efficace combinée à l'acquisition d'informations structurelles de haute qualité grâce au couplage de techniques chromatographiques modernes avec de puissants analyseurs de spectrométrie de masse facilite la détection et l'identification des métabolites dans des échantillons biologiques complexes.
- ◇ La conception correcte de l'étude et la sélection de traitements d'échantillons appropriés sont essentielles pour le développement de stratégies métabolomiques non ciblée.
- ◇ La demande croissante d'un contrôle plus complet du traitement des aliments, en termes d'authenticité, de qualité et de sécurité, peut être satisfaite par la métabolomique non ciblée.
- ◇ L'analyse des composés volatils des échantillons de poisson fumé extraits par DHS-P&T et l'analyse ultérieure par GC-MS ont

permis de développer un modèle de classification pour distinguer les échantillons avec traitement "*Cold smoke*" des autres procédés interdits dans l'Union Européenne.

- ◊ L'application de la métabolomique non ciblée utilisant la nouvelle technique UHPLC-IMS-HRMS a démontré son grand potentiel dans l'identification de biomarqueurs liés à la prise de produits alimentaires et de régimes alimentaires spécifiques, en particulier avec la découverte de biomarqueurs à court terme de la consommation d'oranges.
- ◊ La métabolomique non ciblée peut aider à comprendre les relations complexes entre l'exposition nutritionnelle et l'état de santé, en plus de l'étude de produits alimentaires ou de suppléments potentiellement bénéfiques pour la santé.
- ◊ Le potentiel de l'analyse UHPLC-IMS-HRMS en conjonction avec la métabolomique non ciblée a également été confirmé avec l'élucidation de biomarqueurs liés au traitement au resvératrol et au ptérostilbène chez les rats atteints de stéatose hépatique. Les composés provisoirement élucidés ont démontré le bénéfice potentiel de cette supplémentation pour la santé.
- ◊ Les outils construits autour de la valeur du CCC tels que les bases de données expérimentales du CCS et les outils de prédiction ont facilité l'étape d'identification des composés (bio)marqueurs.
- ◊ L'approche métabolomique non ciblée s'est avérée utile pour déterminer les putatifs signaux chimiques/phéromones dans la communication chimique intra-espèce, en particulier pour la communication mère-enfant chez la souris.
- ◊ La nouvelle procédure d'échantillonnage basée sur le DHS-P&T pour la capture de volatolome à partir de souriceaux vivants et l'analyse ultérieure par GC-MS ont permis l'identification de potentielles phéromones volatiles impliquées dans le comportement maternel.
- ◊ La détection des potentielles signaux chimiques de volatilité faible chez les souriceaux a été rendue possible par une approche métabolomique non ciblée basée sur UHPLC-IMS-HRMS.

Enfin et surtout, tout au long de cette thèse, l'importance de la collaboration multidisciplinaire a été mise en évidence. L'application de méthodologies analytiques avancées pour répondre aux problèmes exposés par les industries alimentaires, les experts en nutrition et les neurobiologistes ont permis d'accroître les connaissances dans ces domaines.

6.2 Future works and prospects

All the work developed in this thesis has been the precedent for the future studies detailed below:

- ◊ Validation of the developed classification model for the differentiation of smoked fish products with fishery products from other companies.
- ◊ Development, validation and application of an analytical targeted method based on GC-EI-Orbitrap for the quantitation and confirmation of the 11 volatile compounds used in the model for the classification of smoked fish. Moreover, additional phenolic compounds and derivatives present in smoked food products will be determined. Different types of on-line volatiles extraction approaches will be evaluated.
- ◊ Design of a study for the determination of medium/long-term biomarkers of orange intake based on a cross-over trial with a wider range of participants and lifestyle.
- ◊ An untargeted lipidomics approach has been carried out to liver samples coming from rats with induced liver steatosis to explore the differences obtained by pterostilbene and resveratrol supplementation in the lipid fingerprint. Statistical data is currently being analysed
- ◊ Apply the untargeted metabolomic analysis to plasma samples collected from the same rats, in order to have a global overview of the circulating metabolites throughout the body and the difference by pterostilbene and resveratrol supplementation.

6.2 Trabajos futuros y perspectivas

Todo el trabajo desarrollado en esta tesis ha sido el precedente para los futuros estudios detallados a continuación:

- ◇ Validación del modelo de clasificación desarrollado para la diferenciación de productos pesqueros ahumados con productos de otras empresas.
- ◇ Desarrollo, validación y aplicación de un método analítico dirigido basado GC-EI-Orbitrap para la cuantificación y confirmación de los 11 compuestos volátiles del modelo para la clasificación de pescado ahumado. Además, también se determinarán otros compuestos fenólicos y derivados presentes en productos alimenticios ahumados. Se evaluarán diferentes sistemas de extracción en línea de compuestos volátiles.
- ◇ Diseñar un estudio para la determinación de biomarcadores de ingesta de naranja a medio/largo plazo basado en un ensayo cruzado (*cross-over*) con una más amplia gama de participantes y distintos estilos de vida.
- ◇ Se ha llevado a cabo un estudio de lipidómica no dirigida con las muestras de hígado procedentes de ratas con esteatosis hepática inducida para explorar las diferencias obtenidas por la suplementación con pterostilbeno y resveratrol en la huella lipídica. Actualmente se están analizando los datos estadísticos.
- ◇ Aplicar el análisis metabolómico no dirigido en muestras de plasma recolectadas de las mismas ratas, con el fin de tener una visión global de los metabolitos circulantes en todo el cuerpo y la diferencias que supone la suplementación con pterostilbeno y resveratrol.

6.2 Travail future et perspectives

Tous les travaux développés dans cette thèse ont été le précédent pour les futures études détaillées ci-dessous :

- ◇ Validation du modèle de classification développé pour différencier les produits de poisson fumé des produits d'autres entreprises.
- ◇ Développement, validation et application d'une méthode analytique dirigée basée sur GC-EI-Orbitrap pour la quantification et la confirmation des 11 composés volatils du modèle de classification des poissons fumés. De plus, d'autres composés phénoliques et dérivés présents dans les produits alimentaires fumés seront également déterminés. Différents systèmes d'extraction en ligne de composés volatils seront évalués.
- ◇ Concevoir une étude pour la détermination de biomarqueurs de la consommation d'oranges à moyen/long terme, basée sur un essai croisé avec un plus large éventail de participants et des modes de vie différents.
- ◇ Une étude lipidomique non ciblée a été menée sur des échantillons de foie de rats atteints de stéatose hépatique induite pour explorer les différences obtenues par la supplémentation en ptérostilbène et resvératrol dans l'empreinte lipidique. Les données statistiques sont en cours d'analyse.
- ◇ Appliquer une analyse métabolomique non ciblée sur des échantillons de plasma prélevés sur les mêmes rats, afin d'avoir une vision globale des métabolites circulant dans l'organisme et des différences qu'apporte la supplémentation en ptérostilbène et resvératrol.

Annex I.
Co-author agreement

Annex I. Co-author agreement

UNIVERSITAT JAUME I
Escola de Doctorat · ED

Tania Portoles Nicolau, como coautor/ coautora doy mi **autorización** a María Leticia Lacalle Bergeron para la presentación de las siguientes publicaciones como parte de su tesis doctoral.

Relación de publicaciones:

- Chromatography hyphenated to high resolution mass spectrometry in untargeted metabolomics for investigation of food (bio)markers.
  María Leticia Lacalle Bergeron; David Izquierdo Sandoval; Juan Vicente Sancho Llopis; Francisco López Benet; Félix Hernández Hernández; Tania Portolés Nicolau.
  TRENDS IN ANALYTICAL CHEMISTRY. 135,pp. 116161. 2021. DOI 10.1016/j.trac.2020.116161

- Gas chromatography-mass spectrometry based untargeted volatolomics for smoked seafood classification.
  María Leticia Lacalle Bergeron; Tania Portolés Nicolau; Carlos Sales Martínez; M. Carmen Corell; Fernando Domínguez; Joaquim Beltrán Arandes; Juan Vicente Sancho Llopis; Félix Hernández Hernández.
  FOOD RESEARCH INTERNATIONAL. 137,pp. 109698. 2020. DOI 10.1016/j.foodres.2020.109698

- Ultra-Performance Liquid Chromatography-Ion Mobility Separation-Quadruple Time-of-Flight MS (UHPLC-IMS-QTOF MS) Metabolomics for Short-Term Biomarker Discovery of Orange Intake: A Randomized, Controlled Crossover Study.
  María Leticia Lacalle Bergeron; Tania Portolés Nicolau; Francisco López Benet; Juan Vicente Sancho Llopis; Carolina Ortega Azorín; Eva M. Asensio; Oscar Coltell Simón; Dolores Corella.
  NUTRIENTS. 12,pp. 1 - 21. 2020. DOI 10.3390/nu12071916 ISSN 2072-6643

- The potential of ion mobility separation in combination with high resolution mass spectrometry for the identification of biomarkers highlighted by untargeted metabolomics: the effects of pterostilbene and resveratrol in liver steatosis, animal model.
  María Leticia Lacalle Bergeron; David Izquierdo Sandoval, Alfredo Fernández Quintela, María Puy Portillo, Juan Vicente Sancho Llopis, Félix Hernández Hernández, Tania Portolés Nicolau.
  Food Chemistry (2021) Submitted

- Novel sampling strategy for alive animal volatolome extraction combined with GC-MS based untargeted metabolomics: Identifying mouse pup pheromones.
  María Leticia Lacalle Bergeron; Rafael Goterris Cerisuelo, Tania Portolés Nicolau, D; Joaquim Beltrán Arandes, Juan Vicente Sancho Llopis; Cinta Navarro Moreno, Fernando Marínez García.
  Talanta 235 (2021) 122786 Contents. DOI 10.1016/j.talanta.2021.122786 ISSN 00399140

- Untargeted metabolomics approach to putative pup pheromones in mice. Part 2: using UHPLC-IMS-QTOF MS of surface body samples to identify low-volatility chemosignals.
  María Leticia Lacalle Bergeron; Rafael Goterris Cerisuelo, Joaquim Beltrán Arandes, Juan Vicente Sancho Llopis, Cinta Navarro Moreno, Fernando Marínez García, Tania Portolés Nicolau.

UNIVERSITAT JAUME I
Escola de Doctorat · ED

Talanta (2021) Submitted

Asimismo, **renuncio** a poder utilizar estas publicaciones como parte de otra tesis doctoral.

Y para que conste firmo el presente documento,

TANIA| PORTOLES |NICOLAU

Firmado digitalmente por TANIA|PORTOLES|NICOLAU
Fecha: 2021.12.07 09:42:29 +01'00'

Lugar, fecha y firma

Todo ello, atendiendo al artículo 28 del Reglamento de los estudios de doctorado de la Universitat Jaume I de Castelló, regulados por el RD 99/2011, en la Universitat Jaume I (Aprobado en la sesión nº 8/2020 del Consejo de Gobierno de 02 /10/2020):

"(...)

4. En el caso de publicaciones conjuntas, todas las personas coautoras deberán manifestar explícitamente su autorización para que la doctoranda o doctorando presente el trabajo como parte de su tesis y la renuncia expresa a presentar este mismo trabajo como parte de otra tesis doctoral. Esta autorización se adjuntará como documentación en el momento del inicio de evaluación de la tesis.

Annex I. Co-author agreement

UNIVERSITAT JAUME I
Escola de Doctorat · ED

JUAN VICENTE SANCHO LLOPIS, como coautor/ coautora doy mi **autorización** a María Leticia Lacalle Bergeron para la presentación de las siguientes publicaciones como parte de su tesis doctoral.

Relación de publicaciones:

- Chromatography hyphenated to high resolution mass spectrometry in untargeted metabolomics for investigation of food (bio)markers.
  María Leticia Lacalle Bergeron; David Izquierdo Sandoval; Juan Vicente Sancho Llopis; Francisco López Benet; Félix Hernández Hernández; Tania Portolés Nicolau.
  TRENDS IN ANALYTICAL CHEMISTRY. 135,pp. 116161. 2021. DOI 10.1016/j.trac.2020.116161
- Gas chromatography-mass spectrometry based untargeted volatolomics for smoked seafood classification.
  María Leticia Lacalle Bergeron; Tania Portolés Nicolau; Carlos Sales Martínez; M. Carmen Corell; Fernando Domínguez; Joaquim Beltrán Arandes; Juan Vicente Sancho Llopis; Félix Hernández Hernández.
  FOOD RESEARCH INTERNATIONAL. 137,pp. 109698. 2020. DOI 10.1016/j.foodres.2020.109698
- Ultra-Performance Liquid Chromatography-Ion Mobility Separation-Quadruple Time-of-Flight MS (UHPLC-IMS-QTOF MS) Metabolomics for Short-Term Biomarker Discovery of Orange Intake: A Randomized, Controlled Crossover Study.
  María Leticia Lacalle Bergeron; Tania Portolés Nicolau; Francisco López Benet; Juan Vicente Sancho Llopis; Carolina Ortega Azorín; Eva M. Asensio; Oscar Coltell Simón; Dolores Corella.
  NUTRIENTS. 12,pp. 1 - 21. 2020. DOI 10.3390/nu12071916 ISSN 2072-6643
- The potential of ion mobility separation in combination with high resolution mass spectrometry for the identification of biomarkers highlighted by untargeted metabolomics: the effects of pterostilbene and resveratrol in liver steatosis, animal model.
  María Leticia Lacalle Bergeron; David Izquierdo Sandoval, Alfredo Fernández Quintela, María Puy Portillo, Juan Vicente Sancho Llopis, Félix Hernández Hernández, Tania Portolés Nicolau.
  Food Chemistry (2021) Submitted
- Novel sampling strategy for alive animal volatolome extraction combined with GC-MS based untargeted metabolomics: Identifying mouse pup pheromones.
  María Leticia Lacalle Bergeron; Rafael Goterris Cerisuelo, Tania Portolés Nicolau, D; Joaquim Beltrán Arandes, Juan Vicente Sancho Llopis; Cinta Navarro Moreno, Fernando Marínez García.
  Talanta 235 (2021) 122786 Contents. DOI 10.1016/j.talanta.2021.122786 ISSN 00399140
- Untargeted metabolomics approach to putative pup pheromones in mice. Part 2: using UHPLC-IMS-QTOF MS of surface body samples to identify low-volatility chemosignals.
  María Leticia Lacalle Bergeron; Rafael Goterris Cerisuelo, Joaquim Beltrán Arandes, Juan Vicente Sancho Llopis, Cinta Navarro Moreno, Fernando Marínez García, Tania Portolés Nicolau.
  Talanta (2021) Submitted

UNIVERSITAT JAUME I
Escola de Doctorat · ED

Asimismo, **renuncio** a poder utilizar estas publicaciones como parte de otra tesis doctoral.

Y para que conste firmo el presente documento,

JUAN VICENTE| SANCHO| LLOPIS

Firmado digitalmente por JUAN VICENTE| SANCHO|LLOPIS
Fecha: 2021.12.05 11:20:50 +01'00'

Castelló, 5 de Diciembre de 2021

Lugar, fecha y firma

Todo ello, atendiendo al artículo 28 del Reglamento de los estudios de doctorado de la Universitat Jaume I de Castelló, regulados por el RD 99/2011, en la Universitat Jaume I (Aprobado en la sesión nº 8/2020 del Consejo de Gobierno de 02 /10/2020):

"(...)

4. En el caso de publicaciones conjuntas, todas las personas coautoras deberán manifestar explícitamente su autorización para que la doctoranda o doctorando presente el trabajo como parte de su tesis y la renuncia expresa a presentar este mismo trabajo como parte de otra tesis doctoral. Esta autorización se adjuntará como documentación en el momento del inicio de evaluación de la tesis.

Annex I. Co-author agreement

UNIVERSITAT JAUME I
Escola de Doctorat · ED

Joaquim Beltrán Arandes.., como coautor/ coautora doy mi **autorización** a María Leticia Lacalle Bergeron para la presentación de las siguientes publicaciones como parte de su tesis doctoral.

Relación de publicaciones:

- Gas chromatography-mass spectrometry based untargeted volatolomics for smoked seafood classification.
  María Leticia Lacalle Bergeron; Tania Portolés Nicolau; Carlos Sales Martínez; M. Carmen Corell; Fernando Domínguez; Joaquim Beltrán Arandes; Juan Vicente Sancho Llopis; Félix Hernández Hernández.
  FOOD RESEARCH INTERNATIONAL. 137,pp. 109698. 2020. DOI 10.1016/j.foodres.2020.109698

- Novel sampling strategy for alive animal volatolome extraction combined with GC-MS based untargeted metabolomics: Identifying mouse pup pheromones.
  María Leticia Lacalle Bergeron; Rafael Goterris Cerisuelo, Tania Portolés Nicolau, D; Joaquim Beltrán Arandes, Juan Vicente Sancho Llopis; Cinta Navarro Moreno, Fernando Marínez García.
  Talanta 235 (2021) 122786 Contents. DOI 10.1016/j.talanta.2021.122786 ISSN 00399140

- Untargeted metabolomics approach to putative pup pheromones in mice. Part 2: using UHPLC-IMS-QTOF MS of surface body samples to identify low-volatility chemosignals.
  María Leticia Lacalle Bergeron; Rafael Goterris Cerisuelo, Joaquim Beltrán Arandes, Juan Vicente Sancho Llopis, Cinta Navarro Moreno, Fernando Marínez García, Tania Portolés Nicolau.
  Talanta (2021) Submitted

Asimismo, **renuncio** a poder utilizar estas publicaciones como parte de otra tesis doctoral.

Y para que conste firmo el presente documento,

JOAQUIN|BELTRAN|ARANDES

Firmado digitalmente por JOAQUIN|BELTRAN|ARANDES
Fecha: 2021.12.09 08:25:36 +01'00'

Lugar, fecha y firma

Todo ello, atendiendo al artículo 28 del Reglamento de los estudios de doctorado de la Universitat Jaume I de Castelló, regulados por el RD 99/2011, en la Universitat Jaume I (Aprobado en la sesión nº 8/2020 del Consejo de Gobierno de 02 /10/2020):

"(...)

4. En el caso de publicaciones conjuntas, todas las personas coautoras deberán manifestar explícitamente su autorización para que la doctoranda o doctorando presente el trabajo como parte de su tesis y la renuncia expresa a presentar este mismo trabajo como parte de otra tesis doctoral. Esta autorización se adjuntará como documentación en el momento del inicio de evaluación de la tesis.

UNIVERSITAT JAUME I
Escola de Doctorat · ED

Félix Hernández Hernández, como coautor doy mi **autorización** a María Leticia Lacalle Bergeron para la presentación de las siguientes publicaciones como parte de su tesis doctoral.

Relación de publicaciones:

- Chromatography hyphenated to high resolution mass spectrometry in untargeted metabolomics for investigation of food (bio)markers.
  María Leticia Lacalle Bergeron; David Izquierdo Sandoval; Juan Vicente Sancho Llopis; Francisco López Benet; Félix Hernández Hernández; Tania Portolés Nicolau.
  TRENDS IN ANALYTICAL CHEMISTRY. 135,pp. 116161. 2021. DOI 10.1016/j.trac.2020.116161

- Gas chromatography-mass spectrometry based untargeted volatolomics for smoked seafood classification.
  María Leticia Lacalle Bergeron; Tania Portolés Nicolau; Carlos Sales Martínez; M. Carmen Corell; Fernando Domínguez; Joaquim Beltrán Arandes; Juan Vicente Sancho Llopis; Félix Hernández Hernández.
  FOOD RESEARCH INTERNATIONAL. 137,pp. 109698. 2020. DOI 10.1016/j.foodres.2020.109698

- The potential of ion mobility separation in combination with high resolution mass spectrometry for the identification of biomarkers highlighted by untargeted metabolomics: the effects of pterostilbene and resveratrol in liver steatosis, animal model.
  María Leticia Lacalle Bergeron; David Izquierdo Sandoval, Alfredo Fernández Quintela, María Puy Portillo, Juan Vicente Sancho Llopis, Félix Hernández Hernández, Tania Portolés Nicolau.
  Food Chemistry (2021) Submitted

Asimismo, **renuncio** a utilizar estas publicaciones como parte de otra tesis doctoral.

Y para que conste firmo el presente documento,

FELIX JAVIER HERNANDEZ HERNANDEZ
Firmado digitalmente por FELIX JAVIER|HERNANDEZ|HERNANDEZ
Fecha: 2021.12.04 18:39:43 +01'00'

Lugar, fecha y firma

Todo ello, atendiendo al artículo 28 del Reglamento de los estudios de doctorado de la Universitat Jaume I de Castelló, regulados por el RD 99/2011, en la Universitat Jaume I (Aprobado en la sesión nº 8/2020 del Consejo de Gobierno de 02 /10/2020):

"(...)

4. En el caso de publicaciones conjuntas, todas las personas coautoras deberán manifestar explícitamente su autorización para que la doctoranda o doctorando presente el trabajo como parte de su tesis y la renuncia expresa a presentar este mismo trabajo como parte de otra tesis doctoral. Esta autorización se adjuntará como documentación en el momento del inicio de evaluación de la tesis.

Annex I. Co-author agreement

UNIVERSITAT JAUME I
Escola de Doctorat · ED

Francisco López Benet..., como coautor/ coautora doy mi **autorización** a María Leticia Lacalle Bergeron para la presentación de las siguientes publicaciones como parte de su tesis doctoral.

Relación de publicaciones:

- Ultra-Performance Liquid Chromatography-Ion Mobility Separation-Quadruple Time-of-Flight MS (UHPLC-IMS-QTOF MS) Metabolomics for Short-Term Biomarker Discovery of Orange Intake: A Randomized, Controlled Crossover Study.
  María Leticia Lacalle Bergeron; Tania Portolés Nicolau; Francisco López Benet; Juan Vicente Sancho Llopis; Carolina Ortega Azorín; Eva M. Asensio; Oscar Coltell Simón; Dolores Corella.
  NUTRIENTS. 12,pp. 1 - 21. 2020. DOI 10.3390/nu12071916 ISSN 2072-6643

- Chromatography hyphenated to high resolution mass spectrometry in untargeted metabolomics for investigation of food (bio)markers.
  María Leticia Lacalle Bergeron; David Izquierdo Sandoval; Juan Vicente Sancho Llopis; Francisco López Benet; Félix Hernández Hernández; Tania Portolés Nicolau.
  TRENDS IN ANALYTICAL CHEMISTRY. 135,pp. 116161. 2021. DOI 10.1016/j.trac.2020.116161

Asimismo, **renuncio** a poder utilizar estas publicaciones como parte de otra tesis doctoral.

Y para que conste firmo el presente documento,

Lugar, fecha y firma

FRANCISCO JOSE|LOPEZ| BENET

Firmado digitalmente por FRANCISCO JOSE|LOPEZ|BENET
Fecha: 2021.12.04 12:44:35 +01'00'

Todo ello, atendiendo al artículo 28 del Reglamento de los estudios de doctorado de la Universitat Jaume I de Castelló, regulados por el RD 99/2011, en la Universitat Jaume I (Aprobado en la sesión nº 8/2020 del Consejo de Gobierno de 02 /10/2020):

UNIVERSITAT JAUME I
Escola de Doctorat · ED

"(...)
4. En el caso de publicaciones conjuntas, todas las personas coautoras deberán manifestar explícitamente su autorización para que la doctoranda o doctorando presente el trabajo como parte de su tesis y la renuncia expresa a presentar este mismo trabajo como parte de otra tesis doctoral. Esta autorización se adjuntará como documentación en el momento del inicio de evaluación de la tesis.

Annex I. Co-author agreement

UNIVERSITAT JAUME I
Escola de Doctorat · ED

David Izquierdo Sandoval ..., como coautor/ coautora doy mi **autorización** a María Leticia Lacalle Bergeron para la presentación de las siguientes publicaciones como parte de su tesis doctoral.

Relación de publicaciones:

- Chromatography hyphenated to high resolution mass spectrometry in untargeted metabolomics for investigation of food (bio)markers.
  María Leticia Lacalle Bergeron; David Izquierdo Sandoval; Juan Vicente Sancho Llopis; Francisco López Benet; Félix Hernández Hernández; Tania Portolés Nicolau.
  TRENDS IN ANALYTICAL CHEMISTRY. 135,pp. 116161. 2021. DOI 10.1016/j.trac.2020.116161

- The potential of ion mobility separation in combination with high resolution mass spectrometry for the identification of biomarkers highlighted by untargeted metabolomics: the effects of pterostilbene and resveratrol in liver steatosis, animal model.
  María Leticia Lacalle Bergeron; David Izquierdo Sandoval, Alfredo Fernández Quintela, María Puy Portillo, Juan Vicente Sancho Llopis, Félix Hernández Hernández, Tania Portolés Nicolau.
  Food Chemistry (2021) Submitted

Asimismo, **renuncio** a poder utilizar estas publicaciones como parte de otra tesis doctoral.

Y para que conste firmo el presente documento,

Lugar, fecha y firma

DAVID| IZQUIERDO| SANDOVAL

Firmado digitalmente por DAVID|IZQUIERDO| SANDOVAL
Fecha: 2021.12.04 12:52:15 +01'00'

Todo ello, atendiendo al artículo 28 del Reglamento de los estudios de doctorado de la Universitat Jaume I de Castelló, regulados por el RD 99/2011, en la Universitat Jaume I (Aprobado en la sesión nº 8/2020 del Consejo de Gobierno de 02 /10/2020):

"(...)

4. En el caso de publicaciones conjuntas, todas las personas coautoras deberán manifestar explícitamente su autorización para que la doctoranda o doctorando presente el trabajo como parte de su tesis y la renuncia expresa a presentar este mismo trabajo como parte de otra tesis doctoral. Esta autorización se adjuntará como documentación en el momento del inicio de evaluación de la tesis.

UNIVERSITAT JAUME I
Escola de Doctorat · ED

Carlos Sales Martinez, como coautor/ coautora doy mi **autorización** a María Leticia Lacalle Bergeron para la presentación de las siguientes publicaciones como parte de su tesis doctoral.

Relación de publicaciones:

- Gas chromatography-mass spectrometry based untargeted volatolomics for smoked seafood classification.
  María Leticia Lacalle Bergeron; Tania Portolés Nicolau; Carlos Sales Martínez, M. Carmen Corell, Fernando Domínguez, Joaquim Beltrán Arandes, Juan Vicente Sancho Llopis; Félix Hernández Hernández.
  Food Research International 137 (2020) 109698. DOI 10.1016/j.foodres.2020.109698
  ISSN 0963-9969

Asimismo, **renuncio** a poder utilizar estas publicaciones como parte de otra tesis doctoral.

Y para que conste firmo el presente documento,

Lugar, fecha y firma

Castellón, 02 de Diciembre de 2021

Todo ello, atendiendo al artículo 28 del Reglamento de los estudios de doctorado de la Universitat Jaume I de Castelló, regulados por el RD 99/2011, en la Universitat Jaume I (Aprobado en la sesión nº 8/2020 del Consejo de Gobierno de 02 /10/2020):

"(...)

4. En el caso de publicaciones conjuntas, todas las personas coautoras deberán manifestar explícitamente su autorización para que la doctoranda o doctorando presente el trabajo como parte de su tesis y la renuncia expresa a presentar este mismo trabajo como parte de otra tesis doctoral. Esta autorización se adjuntará como documentación en el momento del inicio de evaluación de la tesis.

Annex I. Co-author agreement

UNIVERSITAT JAUME I
Escola de Doctorat · ED

M. Carmen Corell como coautor/ coautora doy mi **autorización** a María Leticia Lacalle Bergeron para la presentación de las siguientes publicaciones como parte de su tesis doctoral.

Relación de publicaciones:

- Gas chromatography-mass spectrometry based untargeted volatolomics for smoked seafood classification.
  María Leticia Lacalle Bergeron; Tania Portolés Nicolau; Carlos Sales Martínez, M. Carmen Corell, Fernando Domínguez, Joaquim Beltrán Arandes, Juan Vicente Sancho Llopis; Félix Hernández Hernández.
  Food Research International 137 (2020) 109698. DOI 10.1016/j.foodres.2020.109698
  ISSN 0963-9969

Asimismo, **renuncio** a poder utilizar estas publicaciones como parte de otra tesis doctoral.

Y para que conste firmo el presente documento,

Lugar, fecha y firma
_Las Palmas de Gran Canaria, a 7 de diciembre de 2021

Todo ello, atendiendo al artículo 28 del Reglamento de los estudios de doctorado de la Universitat Jaume I de Castelló, regulados por el RD 99/2011, en la Universitat Jaume I (Aprobado en la sesión nº 8/2020 del Consejo de Gobierno de 02 /10/2020):

"(...)

4. En el caso de publicaciones conjuntas, todas las personas coautoras deberán manifestar explícitamente su autorización para que la doctoranda o doctorando presente el trabajo como parte de su tesis y la renuncia expresa a presentar este mismo trabajo como parte de otra tesis doctoral. Esta autorización se adjuntará como documentación en el momento del inicio de evaluación de la tesis.

UNIVERSITAT JAUME I
Escola de Doctorat · ED

Fernando Domínguez Trenco, como coautor/ coautora doy mi **autorización** a María Leticia Lacalle Bergeron para la presentación de las siguientes publicaciones como parte de su tesis doctoral.

Relación de publicaciones:

- Gas chromatography-mass spectrometry based untargeted volatolomics for smoked seafood classification.
  María Leticia Lacalle Bergeron; Tania Portolés Nicolau; Carlos Sales Martínez, M. Carmen Corell, Fernando Domínguez, Joaquim Beltrán Arandes, Juan Vicente Sancho Llopis; Félix Hernández Hernández.
  Food Research International 137 (2020) 109698. DOI 10.1016/j.foodres.2020.109698
  ISSN 0963-9969

Asimismo, **renuncio** a poder utilizar estas publicaciones como parte de otra tesis doctoral.

Y para que conste firmo el presente documento,

FERNANDO DOMINGUEZ|TRENCO Firmado digitalmente por FERNANDO|DOMINGUEZ|TRENCO Fecha: 2021.12.08 21:11:23 +01'00'

Lugar, fecha y firma

Todo ello, atendiendo al artículo 28 del Reglamento de los estudios de doctorado de la Universitat Jaume I de Castelló, regulados por el RD 99/2011, en la Universitat Jaume I (Aprobado en la sesión nº 8/2020 del Consejo de Gobierno de 02 /10/2020):

"(...)

4. En el caso de publicaciones conjuntas, todas las personas coautoras deberán manifestar explícitamente su autorización para que la doctoranda o doctorando presente el trabajo como parte de su tesis y la renuncia expresa a presentar este mismo trabajo como parte de otra tesis doctoral. Esta autorización se adjuntará como documentación en el momento del inicio de evaluación de la tesis.

Annex I. Co-author agreement

UNIVERSITAT JAUME I
Escola de Doctorat · ED

Dolores Corella Piquer ....., como coautor/ coautora doy mi **autorización** a María Leticia Lacalle Bergeron para la presentación de las siguientes publicaciones como parte de su tesis doctoral.

Relación de publicaciones:

- Ultra-Performance Liquid Chromatography-Ion Mobility Separation-Quadruple Time-of-Flight MS (UHPLC-IMS-QTOF MS) Metabolomics for Short-Term Biomarker Discovery of Orange Intake: A Randomized, Controlled Crossover Study.
  María Leticia Lacalle Bergeron; Tania Portolés Nicolau; Francisco López Benet; Juan Vicente Sancho Llopis; Carolina Ortega Azorín; Eva M. Asensio; Oscar Coltell Simón; Dolores Corella.
  NUTRIENTS. 12,pp. 1 - 21. 2020. DOI 10.3390/nu12071916 ISSN 2072-6643

Asimismo, **renuncio** a poder utilizar estas publicaciones como parte de otra tesis doctoral.

Y para que conste firmo el presente documento,

CORELLA PIQUER DOLORES - 52790127K

Firmado digitalmente por CORELLA PIQUER DOLORES - 52790127K
DN: cn=CORELLA PIQUER DOLORES - 52790127K, c=ES, email=dolores.corella@uv.es
Motivo: Estoy de acuerdo en los términos definidos por firmar este documento
Ubicación: Departamento de Medicina Preventiva y Salud Pública, U. Valencia
Fecha: 2021.07.29 02:57:55 +02'00'

Lugar, fecha y firma

Todo ello, atendiendo al artículo 28 del Reglamento de los estudios de doctorado de la Universitat Jaume I de Castelló, regulados por el RD 99/2011, en la Universitat Jaume I (Aprobado en la sesión nº 8/2020 del Consejo de Gobierno de 02 /10/2020):

"(...)

4. En el caso de publicaciones conjuntas, todas las personas coautoras deberán manifestar explícitamente su autorización para que la doctoranda o doctorando presente el trabajo como parte de su tesis y la renuncia expresa a presentar este mismo trabajo como parte de otra tesis doctoral. Esta autorización se adjuntará como documentación en el momento del inicio de evaluación de la tesis.

UNIVERSITAT JAUME I
Escola de Doctorat · ED

Oscar Coltell Simón ....., como coautor/ coautora doy mi **autorización** a María Leticia Lacalle Bergeron para la presentación de las siguientes publicaciones como parte de su tesis doctoral.

Relación de publicaciones:

- Ultra-Performance Liquid Chromatography-Ion Mobility Separation-Quadruple Time-of-Flight MS (UHPLC-IMS-QTOF MS) Metabolomics for Short-Term Biomarker Discovery of Orange Intake: A Randomized, Controlled Crossover Study.
  María Leticia Lacalle Bergeron; Tania Portolés Nicolau; Francisco López Benet; Juan Vicente Sancho Llopis; Carolina Ortega Azorín; Eva M. Asensio; Oscar Coltell Simón; Dolores Corella.
  NUTRIENTS. 12,pp. 1 - 21. 2020. DOI 10.3390/nu12071916 ISSN 2072-6643

Asimismo, **renuncio** a poder utilizar estas publicaciones como parte de otra tesis doctoral.

Y para que conste firmo el presente documento,

COLTELL SIMON OSCAR - 18921284N

Firmado digitalmente por COLTELL SIMON OSCAR - 18921284N
DN: cn=COLTELL SIMON OSCAR - 18921284N, c=ES, emailvoscar.coltell@uji.es
Motivo: Estoy de acuerdo en los términos definidos por firmar este documento
Ubicación: Departamento de Lenguajes y Sistemas Informáticos, U. Jaume I
Fecha: 2021.07.29 02:59:28 +02'00'

Lugar, fecha y firma

Todo ello, atendiendo al artículo 28 del Reglamento de los estudios de doctorado de la Universitat Jaume I de Castelló, regulados por el RD 99/2011, en la Universitat Jaume I (Aprobado en la sesión nº 8/2020 del Consejo de Gobierno de 02 /10/2020):

"(...)

4. En el caso de publicaciones conjuntas, todas las personas coautoras deberán manifestar explícitamente su autorización para que la doctoranda o doctorando presente el trabajo como parte de su tesis y la renuncia expresa a presentar este mismo trabajo como parte de otra tesis doctoral. Esta autorización se adjuntará como documentación en el momento del inicio de evaluación de la tesis.

Annex I. Co-author agreement

UNIVERSITAT JAUME I
Escola de Doctorat · ED

Carolina Ortega Azorín
....., como coautor/ coautora doy mi **autorización** a María Leticia Lacalle Bergeron para la presentación de las siguientes publicaciones como parte de su tesis doctoral.

Relación de publicaciones:

- Ultra-Performance Liquid Chromatography-Ion Mobility Separation-Quadruple Time-of-Flight MS (UHPLC-IMS-QTOF MS) Metabolomics for Short-Term Biomarker Discovery of Orange Intake: A Randomized, Controlled Crossover Study.
  María Leticia Lacalle Bergeron; Tania Portolés Nicolau; Francisco López Benet; Juan Vicente Sancho Llopis; Carolina Ortega Azorín; Eva M. Asensio; Oscar Coltell Simón; Dolores Corella.
  NUTRIENTS. 12,pp. 1 - 21. 2020. DOI 10.3390/nu12071916 ISSN 2072-6643

Asimismo, **renuncio** a poder utilizar estas publicaciones como parte de otra tesis doctoral.

Y para que conste firmo el presente documento,

CAROLINA| ORTEGA| AZORIN
Firmado digitalmente por CAROLINA| ORTEGA|AZORIN
Fecha: 2021.07.23 14:47:09 +02'00'

Lugar, fecha y firma
Valencia, 23 de Julio de 2021

Todo ello, atendiendo al artículo 28 del Reglamento de los estudios de doctorado de la Universitat Jaume I de Castelló, regulados por el RD 99/2011, en la Universitat Jaume I (Aprobado en la sesión nº 8/2020 del Consejo de Gobierno de 02 /10/2020):

"(...)

4. En el caso de publicaciones conjuntas, todas las personas coautoras deberán manifestar explícitamente su autorización para que la doctoranda o doctorando presente el trabajo como parte de su tesis y la renuncia expresa a presentar este mismo trabajo como parte de otra tesis doctoral. Esta autorización se adjuntará como documentación en el momento del inicio de evaluación de la tesis.

UNIVERSITAT JAUME I
Escola de Doctorat · ED

Eva María Asensio Márquez
....., como coautor/ coautora doy mi **autorización** a María Leticia Lacalle Bergeron para la presentación de las siguientes publicaciones como parte de su tesis doctoral.

Relación de publicaciones:

- Ultra-Performance Liquid Chromatography-Ion Mobility Separation-Quadruple Time-of-Flight MS (UHPLC-IMS-QTOF MS) Metabolomics for Short-Term Biomarker Discovery of Orange Intake: A Randomized, Controlled Crossover Study.
  María Leticia Lacalle Bergeron; Tania Portolés Nicolau; Francisco López Benet; Juan Vicente Sancho Llopis; Carolina Ortega Azorín; Eva M. Asensio; Oscar Coltell Simón; Dolores Corella.
  NUTRIENTS. 12,pp. 1 - 21. 2020. DOI 10.3390/nu12071916 ISSN 2072-6643

Asimismo, **renuncio** a poder utilizar estas publicaciones como parte de otra tesis doctoral.

Y para que conste firmo el presente documento,

EVA MARIA ASENSIO MARQUEZ
Firmado digitalmente por EVA MARIA ASENSIO MARQUEZ
Fecha: 2021.07.29 11:35:11 +02'00'

Lugar, fecha y firma
Valencia, 29 de julio de 2021

Todo ello, atendiendo al artículo 28 del Reglamento de los estudios de doctorado de la Universitat Jaume I de Castelló, regulados por el RD 99/2011, en la Universitat Jaume I (Aprobado en la sesión nº 8/2020 del Consejo de Gobierno de 02 /10/2020):

"(...)

4. En el caso de publicaciones conjuntas, todas las personas coautoras deberán manifestar explícitamente su autorización para que la doctoranda o doctorando presente el trabajo como parte de su tesis y la renuncia expresa a presentar este mismo trabajo como parte de otra tesis doctoral. Esta autorización se adjuntará como documentación en el momento del inicio de evaluación de la tesis.

Annex I. Co-author agreement

UNIVERSITAT JAUME I
Escola de Doctorat · ED

María Puy Portillo Baquedano, como coautora doy mi **autorización** a María Leticia Lacalle Bergeron para la presentación de las siguientes publicaciones como parte de su tesis doctoral.

Relación de publicaciones:

- The potential of ion mobility separation in combination with high resolution mass spectrometry for the identification of biomarkers highlighted by untargeted metabolomics: the effects of pterostilbene and resveratrol in liver steatosis, animal model.
  María Leticia Lacalle Bergeron; David Izquierdo Sandoval, Alfredo Fernández Quintela, María Puy Portillo, Juan Vicente Sancho Llopis, Félix Hernández Hernández, Tania Portolés Nicolau.
  Food Chemistry (2021) Submitted

Asimismo, **renuncio** a poder utilizar estas publicaciones como parte de otra tesis doctoral.

Y para que conste firmo el presente documento,

Vitoria-Gasteiz, 4 de Diciembre de 2021

Todo ello, atendiendo al artículo 28 del Reglamento de los estudios de doctorado de la Universitat Jaume I de Castelló, regulados por el RD 99/2011, en la Universitat Jaume I (Aprobado en la sesión nº 8/2020 del Consejo de Gobierno de 02 /10/2020):

"(...)

4. En el caso de publicaciones conjuntas, todas las personas coautoras deberán manifestar explícitamente su autorización para que la doctoranda o doctorando presente el trabajo como parte de su tesis y la renuncia expresa a presentar este mismo trabajo como parte de otra tesis doctoral. Esta autorización se adjuntará como documentación en el momento del inicio de evaluación de la tesis.

UNIVERSITAT JAUME I
Escola de Doctorat · ED

Dr. Alfredo Fernández Quintela, como coautor doy mi **autorización** a María Leticia Lacalle Bergeron para la presentación de las siguientes publicaciones como parte de su tesis doctoral.

Relación de publicaciones:

- The potential of ion mobility separation in combination with high resolution mass spectrometry for the identification of biomarkers highlighted by untargeted metabolomics: the effects of pterostilbene and resveratrol in liver steatosis, animal model.
  María Leticia Lacalle Bergeron; David Izquierdo Sandoval, Alfredo Fernández Quintela, María Puy Portillo, Juan Vicente Sancho Llopis, Félix Hernández Hernández, Tania Portolés Nicolau.
  Food Chemistry (2021) Submitted

Asimismo, **renuncio** a poder utilizar estas publicaciones como parte de otra tesis doctoral.

Y para que conste firmo el presente documento,

Alfredo Fernández Quintela
Firmado digitalmente por Alfredo Fernández Quintela
Fecha: 2021.12.03 16:55:27 +01'00'

Lugar, fecha y firma

Todo ello, atendiendo al artículo 28 del Reglamento de los estudios de doctorado de la Universitat Jaume I de Castelló, regulados por el RD 99/2011, en la Universitat Jaume I (Aprobado en la sesión nº 8/2020 del Consejo de Gobierno de 02 /10/2020):

"(...)

4. En el caso de publicaciones conjuntas, todas las personas coautoras deberán manifestar explícitamente su autorización para que la doctoranda o doctorando presente el trabajo como parte de su tesis y la renuncia expresa a presentar este mismo trabajo como parte de otra tesis doctoral. Esta autorización se adjuntará como documentación en el momento del inicio de evaluación de la tesis.

Annex I. Co-author agreement

UNIVERSITAT JAUME I
Escola de Doctorat · ED

Ferran Martinez García, como coautor/ coautora doy mi **autorización** a María Leticia Lacalle Bergeron para la presentación de las siguientes publicaciones como parte de su tesis doctoral.

Relación de publicaciones:

- Novel sampling strategy for alive animal volatolome extraction combined with GC-MS based untargeted metabolomics: Identifying mouse pup pheromones.
  María Leticia Lacalle Bergeron; Rafael Goterris Cerisuelo, Tania Portolés Nicolau, D; Joaquim Beltrán Arandes, Juan Vicente Sancho Llopis; Cinta Navarro Moreno, Fernando Marínez García.
  Talanta 235 (2021) 122786 Contents. DOI 10.1016/j.talanta.2021.122786 ISSN 00399140

- Untargeted metabolomics approach to putative pup pheromones in mice. Part 2: using UHPLC-IMS-QTOF MS of surface body samples to identify low-volatility chemosignals.
  María Leticia Lacalle Bergeron; Rafael Goterris Cerisuelo, Joaquim Beltrán Arandes, Juan Vicente Sancho Llopis, Cinta Navarro Moreno, Fernando Marínez García, Tania Portolés Nicolau.
  Talanta (2021) Submitted

Asimismo, **renuncio** a poder utilizar estas publicaciones como parte de otra tesis doctoral.

Y para que conste firmo el presente documento,

Firmado electronicamente por Fernando Martinez Garcia el 09/12/2021, DNI 1892984P, con certificado electrónico emitido por la AVVC (España)
Lugar, fecha y firma

Todo ello, atendiendo al artículo 28 del Reglamento de los estudios de doctorado de la Universitat Jaume I de Castelló, regulados por el RD 99/2011, en la Universitat Jaume I (Aprobado en la sesión nº 8/2020 del Consejo de Gobierno de 02 /10/2020):

"(...)

4. En el caso de publicaciones conjuntas, todas las personas coautoras deberán manifestar explícitamente su autorización para que la doctoranda o doctorando presente el trabajo como parte de su tesis y la renuncia expresa a presentar este mismo trabajo como parte de otra tesis doctoral. Esta autorización se adjuntará como documentación en el momento del inicio de evaluación de la tesis.

UNIVERSITAT JAUME I
Escola de Doctorat · ED

Rafael Goterris Cerisuelo, como coautor/ coautora doy mi **autorización** a María Leticia Lacalle Bergeron para la presentación de las siguientes publicaciones como parte de su tesis doctoral.

Relación de publicaciones:

- Novel sampling strategy for alive animal volatolome extraction combined with GC-MS based untargeted metabolomics: Identifying mouse pup pheromones.
  María Leticia Lacalle Bergeron; Rafael Goterris Cerisuelo, Tania Portolés Nicolau, D; Joaquim Beltrán Arandes, Juan Vicente Sancho Llopis; Cinta Navarro Moreno, Fernando Marínez García.
  Talanta 235 (2021) 122786 Contents. DOI 10.1016/j.talanta.2021.122786 ISSN 00399140

- Untargeted metabolomics approach to putative pup pheromones in mice. Part 2: using UHPLC-IMS-QTOF MS of surface body samples to identify low-volatility chemosignals.
  María Leticia Lacalle Bergeron; Rafael Goterris Cerisuelo, Joaquim Beltrán Arandes, Juan Vicente Sancho Llopis, Cinta Navarro Moreno, Fernando Marínez García, Tania Portolés Nicolau.
  Talanta (2021) Submitted

Asimismo, **renuncio** a poder utilizar estas publicaciones como parte de otra tesis doctoral.

Y para que conste firmo el presente documento,

Firmado por Rafael Goterris Cerisuelo el 09/12/21 en Castellón de la plana.

Lugar, fecha y firma

Todo ello, atendiendo al artículo 28 del Reglamento de los estudios de doctorado de la Universitat Jaume I de Castelló, regulados por el RD 99/2011, en la Universitat Jaume I (Aprobado en la sesión nº 8/2020 del Consejo de Gobierno de 02 /10/2020):

"(...)

4. En el caso de publicaciones conjuntas, todas las personas coautoras deberán manifestar explícitamente su autorización para que la doctoranda o doctorando presente el trabajo como parte de su tesis y la renuncia expresa a presentar este mismo trabajo como parte de otra tesis doctoral. Esta autorización se adjuntará como documentación en el momento del inicio de evaluación de la tesis.

Annex I. Co-author agreement

UNIVERSITAT JAUME I
Escola de Doctorat · ED

CINTA NAVARRO MORENO como coautor/ coautora doy mi **autorización** a María Leticia Lacalle Bergeron para la presentación de las siguientes publicaciones como parte de su tesis doctoral.

Relación de publicaciones:

- Novel sampling strategy for alive animal volatolome extraction combined with GC-MS based untargeted metabolomics: Identifying mouse pup pheromones.
  María Leticia Lacalle Bergeron; Rafael Goterris Cerisuelo, Tania Portolés Nicolau, D; Joaquim Beltrán Arandes, Juan Vicente Sancho Llopis; Cinta Navarro Moreno, Fernando Marínez García.
  Talanta 235 (2021) 122786 Contents. DOI 10.1016/j.talanta.2021.122786 ISSN 00399140

- Untargeted metabolomics approach to putative pup pheromones in mice. Part 2: using UHPLC-IMS-QTOF MS of surface body samples to identify low-volatility chemosignals.
  María Leticia Lacalle Bergeron; Rafael Goterris Cerisuelo, Joaquim Beltrán Arandes, Juan Vicente Sancho Llopis, Cinta Navarro Moreno, Fernando Marínez García, Tania Portolés Nicolau.
  Talanta (2021) Submitted

Asimismo, **renuncio** a poder utilizar estas publicaciones como parte de otra tesis doctoral.

Y para que conste firmo el presente documento,

Lugar, fecha y firma VALENCIA, 9 DE DICIEMBRE DE 2021

Todo ello, atendiendo al artículo 28 del Reglamento de los estudios de doctorado de la Universitat Jaume I de Castelló, regulados por el RD 99/2011, en la Universitat Jaume I (Aprobado en la sesión nº 8/2020 del Consejo de Gobierno de 02 /10/2020):

"(...)

4. En el caso de publicaciones conjuntas, todas las personas coautoras deberán manifestar explícitamente su autorización para que la doctoranda o doctorando presente el trabajo como parte de su tesis y la renuncia expresa a presentar este mismo trabajo como parte de otra tesis doctoral. Esta autorización se adjuntará como documentación en el momento del inicio de evaluación de la tesis.